

**BEST AVAILABLE COPY**

Docket No.: R2184.0106/P106  
(PATENT)

#3

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of:  
Kouichi Narahara

Application No.: Not Yet Assigned

Group Art Unit: N/A

Filed: July 6, 2001

Examiner: Not Yet Assigned



For: DOCUMENT INFORMATION  
PROCESSING DEVICE THAT ACHIEVES  
EFFICIENT UNDERSTANDING OF  
CONTENTS OF DOCUMENT  
INFORMATION

**CLAIM FOR PRIORITY AND SUBMISSION OF DOCUMENTS**

Commissioner for Patents  
Washington, DC 20231

Dear Sir:

Applicant hereby claims priority under 35 U.S.C. 119 based on the following prior foreign application filed in the following foreign country on the date indicated:

Country	Application No.	Date
Japan	2000-214394	July 14, 2000

In support of this claim, a certified copy of said original foreign application is filed herewith.

Dated: July 6, 2001

Respectfully submitted,

By Mark J. Thronson

Mark J. Thronson  
Registration No.: 33,082  
DICKSTEIN SHAPIRO MORIN &  
OSHINSKY LLP  
2101 L Street NW  
Washington, DC 20037-1526  
(202) 775-4742  
Attorneys for Applicant

日本特許庁  
JAPAN PATENT OFFICE

09/900186 U.S. PRO  
09/900186  
09/900186

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日  
Date of Application:

2000年 7月14日

出願番号  
Application Number:

特願2000-214394

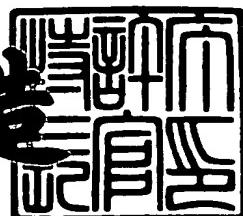
出願人  
Applicant(s):

株式会社リコー

2001年 5月11日

特許庁長官  
Commissioner,  
Japan Patent Office

及川耕造



出証番号 出証特2001-3038348

【書類名】 特許願  
【整理番号】 0004325  
【提出日】 平成12年 7月14日  
【あて先】 特許庁長官 殿  
【国際特許分類】 G06F 12/00  
G06F 13/00  
【発明の名称】 文書情報処理方法、文書情報処理装置及び記録媒体  
【請求項の数】 48  
【発明者】  
【住所又は居所】 東京都大田区中馬込1丁目3番6号 株式会社 リコー  
内  
【氏名】 楢原 孝一  
【特許出願人】  
【識別番号】 000006747  
【氏名又は名称】 株式会社 リコー  
【代表者】 桜井 正光  
【代理人】  
【識別番号】 100073760  
【弁理士】  
【氏名又は名称】 鈴木 誠  
【選任した代理人】  
【識別番号】 100097652  
【弁理士】  
【氏名又は名称】 大浦 一仁  
【手数料の表示】  
【予納台帳番号】 011800  
【納付金額】 21,000円  
【提出物件の目録】  
【物件名】 明細書 1

特2000-214394

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809191

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書情報処理方法、文書情報処理装置及び記録媒体

【特許請求の範囲】

【請求項1】 複数の要素からなる文書情報から印刷装置、表示装置などの画像出力装置又はファクス装置などの画像伝送装置に出力するための出力文書を生成する文書情報処理方法であって、

前記文書情報の各要素の重要度を評価し、前記文書情報の要素を重要度の高い順に選択して出力文書上に配置することを特徴とする文書情報処理方法。

【請求項2】 前記文書情報中に記述された重要度定義情報に基づいて前記文書情報の各要素の重要度を評価することを特徴とする請求項1記載の文書情報処理方法。

【請求項3】 所定の重要度評価基準に基づいて前記文書情報の各要素の重要度を評価することを特徴とする請求項1記載の文書情報処理方法。

【請求項4】 予め指定された前記出力文書のページサイズ及びページ数に基づいて、前記出力文書上に配置される前記文書情報の要素を制限することを特徴とする請求項2又は3記載の文書情報処理方法。

【請求項5】 前記文書情報の選択された要素により占有される前記出力文書上のスペースを積算したスペースが、前記出力文書の指定されたページサイズ及びページ数によって決まるスペース上限以下となるように、前記出力文書上に配置される前記文書情報の要素を制限することを特徴とする請求項4記載の文書情報処理方法。

【請求項6】 前記文書情報の選択された要素により占有される前記出力文書上のスペースを積算したスペースが、前記スペース上限を越えるまで、前記文書情報の要素の選択を行い、最後に選択された要素を前記出力文書から削除することを特徴とする請求項5記載の文書情報処理方法。

【請求項7】 前記文書情報の選択された要素により占有される前記出力文書上のスペースを積算したスペースが、前記出力文書の前記指定されたページサイズ及びページ数によって決まるスペース上限を越えるまで、前記文書情報の要素の選択を行い、前記選択された要素の占有スペースの積算スペースが前記スペ

ース上限以下となるように前記選択された要素中の少なくとも一部の要素を縮小して前記出力文書上に配置することを特徴とする請求項4記載の文書情報処理方法。

【請求項8】 文書情報ソースより、複数の要素からなる文書情報を入力する文書情報入力手段と、

前記文書情報入力手段による入力文書情報の各要素の重要度を評価する重要度評価手段と、

前記入力文書情報の要素を、前記重要度評価手段により評価された重要度の高い順に選択し、選択した要素を選択順に配置した出力文書を生成する文書情報処理手段と、

前記文書情報処理手段により生成された前記出力文書を印刷装置、表示装置などの画像出力装置又はファクス装置などの画像伝送装置へ出力する手段とを有することを特徴とする文書情報処理装置。

【請求項9】 前記重要度評価手段は前記入力文書情報中に記述された重要度定義情報に基づいて前記入力文書情報の各要素の重要度を評価することを特徴とする請求項8記載の文書情報処理装置。

【請求項10】 前記重要度評価手段は所定の重要度評価基準に基づいて前記入力文書情報の各要素の重要度を評価することを特徴とする請求項8記載の文書情報処理装置。

【請求項11】 前記文書情報処理手段は、出力制約条件情報により指定された前記出力文書のページサイズ及びページ数に基づいて、前記出力文書に配置される前記入力文書情報の要素を制限することを特徴とする請求項9又は10記載の文書情報処理装置。

【請求項12】 前記文書情報処理手段は、選択した前記入力文書情報の要素により占有される前記出力文書上のスペースを積算し、積算したスペースが、前記出力文書の指定されたページサイズ及びページ数によって決まるスペース上限以下となるように、前記出力文書に配置される前記入力文書情報の要素を制限することを特徴とする請求項11記載の文書情報処理装置。

【請求項13】 前記文書情報処理手段は、選択した前記文書情報の要素に

より占有される前記出力文書上のスペースを積算し、積算したスペースが前記スペース上限を越えるまで前記入力文書情報の要素の選択を行い、最後に選択した要素を前記出力文書から削除することを特徴とする請求項12記載の文書情報処理装置。

【請求項14】 前記文書情報処理手段は、選択した前記文書情報の要素により占有される前記出力文書上のスペースを積算し、積算したスペースが前記出力文書の指定されたページサイズ及びページ数によって決まるスペース上限を越えるまで前記入力文書情報の要素の選択を行い、前記選択した要素中の少なくとも一部の要素を縮小することを特徴とする請求項11記載の文書情報処理装置。

【請求項15】 前記文書情報ソースを有することを特徴とする請求項9、10又は11記載の文書情報処理装置。

【請求項16】 前記文書情報ソースとネットワークを介して接続されることを特徴とする請求項9、10又は11記載の文書情報処理装置。

【請求項17】 前記画像出力装置又は前記画像伝送装置を有することを特徴とする請求項9、10又は11記載の文書情報処理装置。

【請求項18】 前記画像出力装置又は前記画像伝送装置とネットワークを介して接続されることを特徴とする請求項9、10又は11記載の文書情報処理装置。

【請求項19】 前記出力制約情報を設定する手段を有することを特徴とする請求項11記載の文書情報処理装置。

【請求項20】 前記出力制約情報はネットワークを介して設定されることを特徴とする請求項11記載の文書情報処理装置。

【請求項21】 複数の要素からなる文書情報の処理方法であって、前記文書情報の各要素の重要度を評価し、前記文書情報の各要素にその重要度に応じた操作を施すことにより、前記文書情報の情報量を削減することを特徴とする文書情報処理方法。

【請求項22】 前記文書情報中に記述された重要度定義情報に基づいて、前記文書情報の各要素の重要度を評価することを特徴とする請求項21記載の文書情報処理方法。

【請求項23】 所定の重要度評価基準に基づいて、前記文書情報の各要素の重要度を評価することを特徴とする請求項21記載の文書情報処理方法。

【請求項24】 特定の重要度レベルより低い重要度の要素を削除することを特徴とする請求項21又は22記載の文書情報処理方法。

【請求項25】 要素の属性によって前記特定の重要度レベルが異なることを特徴とする請求項24記載の文書情報処理方法。

【請求項26】 テキスト属性の要素に関する前記特定の重要度レベルより、非テキスト属性の要素に関する前記特定の重要度レベルが高いことを特徴とする請求項25記載の文書情報処理方法。

【請求項27】 テキスト属性の要素を削除対象から除外し、非テキスト属性の要素のみ削除対象とすることを特徴とする請求項24記載の文書情報処理方法。

【請求項28】 非テキスト属性の要素を、その重要度に応じた圧縮方法で圧縮することを特徴とする請求項22又は23記載の文書情報処理方法。

【請求項29】 非テキスト属性の要素を、その重要度に応じた圧縮率で圧縮することを特徴とする請求項22又は23記載の文書情報処理方法。

【請求項30】 第1の重要度レベルより重要度が低いテキスト属性の要素を削除し、第2の重要度レベルより重要度が低い非テキスト属性の要素を圧縮することを特徴とする請求項22又は23記載の文書情報処理方法。

【請求項31】 文書情報ソースより、複数の要素からなる文書情報を入力する文書情報入力手段と、

前記文書情報入力手段による入力文書情報の各要素の重要度を評価する重要度評価手段と、

前記入力文書情報の各要素に、前記重要度評価手段により評価された重要度に応じた操作を施すことにより前記文書情報の情報量を削減する文書情報処理手段と、

前記文書情報処理手段により情報量を削減された文書情報を文書情報保存装置に出力する手段とを有することを特徴とする文書情報処理装置。

【請求項32】 前記重要度評価手段は、前記文書情報中に記述された重要

度定義情報に基づいて前記文書情報の各要素の重要度を評価することを特徴とする請求項31記載の文書情報処理装置。

【請求項33】 前記重要度評価手段は、所定の重要度評価基準に基づいて前記文書情報の各要素の重要度を評価することを特徴とする請求項31記載の文書情報処理装置。

【請求項34】 前記文書情報処理手段は、特定の重要度レベルより低い重要度の要素を削除することを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項35】 要素の属性によって前記特定の重要度レベルが異なることを特徴とする請求項34記載の文書情報処理装置。

【請求項36】 テキスト属性の要素に関する前記特定の重要度レベルより、非テキスト属性の要素に関する前記特定の重要度レベルが高いことを特徴とする請求項35記載の文書情報処理装置。

【請求項37】 テキスト属性の要素は削除対象から除外し、非テキスト属性の要素のみ削除対象とすることを特徴とする請求項34記載の文書情報処理装置。

【請求項38】 前記文書情報処理手段は、非テキスト属性の要素を、その重要度に応じた圧縮方法で圧縮することを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項39】 前記文書情報処理手段は、非テキスト属性の要素を、その重要度に応じた圧縮率で圧縮することを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項40】 前記文書情報処理手段は、第1の重要度レベルより重要度が低いテキスト属性の要素を削除し、第2の重要度レベルより重要度が低い非テキスト属性の要素を圧縮することを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項41】 前記文書情報ソースを有することを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項42】 前記文書情報ソースとネットワークを介して接続されるこ

とを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項43】 前記文書情報保存装置を有することを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項44】 前記文書情報保存装置とネットワークを介して接続されることを特徴とする請求項32又は33記載の文書情報処理装置。

【請求項45】 コンピュータが読み取り可能な記録媒体であって、請求項1乃至7のいずれか1項記載の文書処理方法をコンピュータで実施するためのプログラムが記録されたことを特徴とする記録媒体。

【請求項46】 コンピュータが読み取り可能な記録媒体であって、請求項8乃至20のいずれか1項記載の文書情報処理装置をコンピュータで実現するためのプログラムが記録されたことを特徴とする記録媒体。

【請求項47】 コンピュータが読み取り可能な記録媒体であって、請求項21乃至30のいずれか1項記載の文書情報処理方法をコンピュータで実施するためのプログラムが記録されたことを特徴とする記録媒体。

【請求項48】 コンピュータが読み取り可能な記録媒体であって、請求項31乃至44のいずれか1項記載の文書情報処理装置をコンピュータで実現するためのプログラムが記録されたことを特徴とする記録媒体。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

HTML (Hypertext Markup Language) や XML (Extensible Markup Language) などの構造化文書記述言語で記述された構造化文書情報のような、複数の要素からなる文書情報を処理する装置もしくはシステムに関する。

##### 【0002】

##### 【従来の技術】

HTMLで作成された文書を閲覧するためのWWW(world wide web)ブラウザ・ソフトの中には、ダウンロード時間の短縮のために、米国ネットスケープ・コミュニケーションズ社のNetscape Communicator（登録商標）のように、テキスト情報のみをダウンロードして、それをほぼ元のレイアウトで表示し、画像情報

はダウンロードせず小さなアイコンに置き換えて表示する機能を有するものがある。

【0003】

また、同様の目的で、HTML文書中の特定形式の情報（例えばテキスト情報、GIF形式の画像情報）だけをダウンロードし、それ以外の形式の情報（例えばJPEG形式、BMP形式、MPEG形式の画像情報）や、広告情報、データ量が大きすぎる画像情報などはダウンロードせず、ダウンロードした情報だけを元のレイアウトに従って印刷するネットワークプリンタが特開平11-203100号公報に開示されている。

【0004】

また、WWWブラウザ・ソフト、エディタなどの各種アプリケーション・ソフト、プリンタドライバ・ソフトなどは、ページ指定や範囲指定によって、文書の特定のページや特定ページの一部だけを印刷させる機能を持っている。

【0005】

【発明が解決しようとする課題】

HTMLやXMLなどで記述された構造化文書情報のような、複数の要素からなる文書情報の内容を把握する目的のためには、文書情報の全体を印刷することは必ずしも望ましくはない。一般に、文書情報の内容把握のためには、全ての要素が重要であるというわけではなく、要素によって、その重要度に高低がある。したがって、重要度の高い要素だけを印刷したほうが内容把握を効率的に行うことができ、また、印刷用紙を節約できる。

【0006】

しかし、WWWブラウザ・ソフトなどで文書情報を表示し、重要度の高い要素が含まれているページを指定したり、重要度の高い要素の範囲を指定して印刷を行うことは、操作に手間が掛かるうえに、必ずしも効率的な内容把握や印刷用紙の節約を可能にするものではない。

【0007】

例えば、5ページの文書情報の1ページ目と5ページ目に重要度の高い要素が含まれている場合に、1ページ目と5ページ目だけをページ指定して印刷すれば

印刷用紙の消費枚数を2枚に減らすことができる。しかし、全てのページに重要度の高い要素が分散している場合には、そのようなページ指定によって印刷用紙を節約することはできない。また、いずれの場合であっても、印刷された各ページに重要度の低い要素が含まれていると、効率的な内容把握の妨げになる。

#### 【0008】

ページ指定ではなく、重要度の高い要素の範囲を指定して印刷するならば、重要度の低い要素は印刷されないため、内容把握の効率化のためには有利である。しかし、操作に手間取るうえに、1ページ中に重要度の高い要素が複数含まれている場合には、それぞれの要素を範囲指定して印刷すると、指定した要素がそれぞれ1枚の印刷用紙に印刷されるため、逆に印刷用紙の消費枚数が増加する可能性もある。

#### 【0009】

したがって、本発明の1つの目的は、複数の要素からなる文書情報を印刷する場合に、文書情報の内容の効率的な把握と印刷用紙の節約を可能にする文書情報処理方法及び装置を提供することにある。

#### 【0010】

なお、この目的を前記ネットワークプリンタで達成することは困難である。前記ネットワークプリンタでは、一部の情報がダウンロードされないとはいっても、文書情報は元のレイアウトのまま印刷されるため印刷用紙の節約を期待できないからである。また、要素の情報形式と重要度とは必ずしも対応しないため、重要度の低い要素が印刷されることも、重要度の高い要素が印刷されないこともありますからである。

#### 【0011】

文書情報を表示装置で表示する場合でも、重要度の高い要素だけが表示されるか、少なくとも重要度の高い要素が優先的に表示されると効率的な内容把握が可能である。しかし、前記Netscape CommunicatorのようなWWWブラウザ・ソフトを利用しても、そのような表示をさせることは不可能である。

#### 【0012】

したがって、本発明のもう1つの目的は、表示装置の画面上で、複数の要素か

らなる文書情報の効率的な内容把握を可能にする文書情報処理方法及び装置を提供することにある。

#### 【0013】

文書情報をファクス装置などの画像伝送装置で送信させる場合にも、印刷装置や表示装置に文書情報を出力する場合と同様のことが言える。すなわち、文書情報の重要な要素だけを出力するほうが、受信側で記録紙を節約することができ、また、送信時間を短縮できる。

#### 【0014】

したがって、本発明の1つの目的は、複数の要素からなる文書情報をファクス装置などの画像伝送装置で送信させる場合に、文書情報の内容の効率的な把握、記録紙の節約及び送信時間の短縮を可能にする文書情報処理方法及び装置を提供することにある。

#### 【0015】

XML文書情報のような画像や音声などを含む文書情報は、全体の情報量が大きくなりやすい。このような文書情報では、その保存に必要な記憶容量や伝送に要する時間を減らすために、保存又は伝送に先立って情報量を削減する必要性が高い。しかし、情報量の削減によって文書情報の内容把握に必要な情報が失われたのでは好ましくない。

#### 【0016】

よって、本発明のもう1つの目的は、複数の要素からなる文書情報の内容把握に必要な情報をできるだけ保存しつつ、全体の情報量を削減するための文書情報処理方法及び装置を提供することにある。

#### 【0017】

##### 【課題を解決するための手段】

請求項1乃至7記載の発明は、複数の要素からなる文書情報を印刷、表示又は送信する場合に、文書情報の内容の効率的な把握、印刷用紙もしくは記録紙の節約、さらには送信時間の短縮を可能にする文書情報処理方法に係るもので、その特徴は以下の通りである。

#### 【0018】

請求項1記載の発明の特徴は、文書情報の各要素の重要度を評価し、前記文書情報の要素を重要度の高い順に選択して、画像出力装置又は画像伝送装置への出力文書上に配置することである。

【0019】

請求項2記載の発明の特徴は、請求項1記載の発明による文書情報処理方法において、前記文書情報中に記述された重要度定義情報に基づいて前記文書情報の各要素の重要度を評価することである。

【0020】

請求項3記載の発明の特徴は、請求項1記載の発明による文書情報処理方法において、所定の重要度評価基準に基づいて前記文書情報の各要素の重要度を評価することである。

【0021】

請求項4記載の発明の特徴は、請求項2又は3記載の発明による文書情報処理方法において、予め指定された前記出力文書のページサイズ及びページ数に基づいて、前記出力文書上に配置される前記文書情報の要素を制限することである。

【0022】

請求項5記載の発明の特徴は、請求項4記載の発明による文書情報処理方法において、前記文書情報の選択された要素により占有される前記出力文書上のスペースを積算したスペースが、前記出力文書の指定されたページサイズ及びページ数によって決まるスペース上限以下となるように、前記出力文書上に配置される前記文書情報の要素を制限することである。

【0023】

請求項6記載の発明の特徴は、請求項5記載の発明による文書情報処理方法において、前記文書情報の選択された要素により占有される前記出力文書上のスペースを積算したスペースが前記スペース上限を越えるまで、前記文書情報の要素の選択を行い、最後に選択された要素を前記出力文書から削除することである。

【0024】

請求項7記載の発明の特徴は、請求項5記載の発明による文書情報処理方法において、前記文書情報の選択された要素により占有される前記出力文書上のスペ

ースを積算したスペースが、前記出力文書の前記指定されたページサイズ及びページ数によって決まるスペース上限を越えるまで、前記文書情報の要素の選択を行い、前記選択された要素の占有スペースの積算スペースが前記スペース上限以下となるように前記選択された要素中の少なくとも一部の要素を縮小して前記出力文書上に配置することである。

#### 【0025】

請求項8乃至20記載の発明は、複数の要素からなる文書情報を印刷、表示又は送信する場合に、文書情報の内容の効率的な把握、印刷用紙もしくは記録紙の節約、さらには送信時間の短縮を可能にする文書情報処理装置に係るもので、その特徴は以下の通りである。

#### 【0026】

請求項8記載の発明の特徴は、文書情報ソースより、複数の要素からなる文書情報を入力する文書情報入力手段と、前記文書情報入力手段による入力文書情報の各要素の重要度を評価する重要度評価手段と、前記入力文書情報の要素を、前記重要度評価手段により評価された重要度の高い順に選択し、選択した要素を選択順に配置した出力文書を生成する文書情報処理手段と、前記文書情報処理手段により生成された前記出力文書を印刷装置、表示装置などの画像出力装置又はファクス装置などの画像伝送装置へ出力する手段とを有することである。

#### 【0027】

請求項9記載の発明の特徴は、請求項8記載の発明による文書情報処理装置において、前記重要度評価手段が、前記入力文書情報中に記述された重要度定義情報に基づいて前記入力文書情報の各要素の重要度を評価することである。

#### 【0028】

請求項10記載の発明の特徴は、請求項8記載の発明による文書情報処理装置において、前記重要度評価手段が、所定の重要度評価基準に基づいて前記入力文書情報の各要素の重要度を評価することである。

#### 【0029】

請求項11記載の発明の特徴は、請求項9又は10記載の発明による文書情報処理装置において、前記文書情報処理手段が、出力制約条件情報により指定され

た前記出力文書のページサイズ及びページ数に基づいて、前記出力文書に配置される前記入力文書情報の要素を制限することである。

【0030】

請求項12記載の発明の特徴は、請求項11記載の発明による文書情報処理装置において、前記文書情報処理手段が、選択した前記入力文書情報の要素により占有される前記出力文書上のスペースを積算し、積算したスペースが、前記出力文書の指定されたページサイズ及びページ数によって決まるスペース上限以下となるように、前記出力文書に配置される前記入力文書情報の要素を制限することである。

【0031】

請求項13記載の発明の特徴は、請求項12記載の発明による文書情報処理装置において、前記文書情報処理手段が、選択した前記文書情報の要素により占有される前記出力文書上のスペースを積算し、積算したスペースが前記スペース上限を越えるまで前記入力文書情報の要素の選択を行い、最後に選択した要素を前記出力文書から削除することである。

【0032】

請求項14記載の発明の特徴は、請求項11記載の発明による文書情報処理装置において、前記文書情報処理手段が、選択した前記文書情報の要素により占有される前記出力文書上のスペースを積算し、積算したスペースが前記出力文書の指定されたページサイズ及びページ数によって決まるスペース上限を越えるまで前記入力文書情報の要素の選択を行い、前記選択した要素中の少なくとも一部の要素を縮小することである。

【0033】

請求項15記載の発明の特徴は、請求項9、10又は11記載の発明による文書情報処理装置であって、前記文書情報ソースを有することである。

【0034】

請求項16記載の発明の特徴は、請求項9、10又は11記載の発明による文書情報処理装置であって、前記文書情報ソースとネットワークを介して接続されることである。

【0035】

請求項17記載の発明の特徴は、請求項9、10又は11記載の発明による文書情報処理装置であって、前記画像出力装置又は前記画像伝送装置を有することである。

【0036】

請求項18記載の発明の特徴は、請求項9、10又は11記載の発明による文書情報処理装置であって、前記画像出力装置又は前記画像伝送装置とネットワークを介して接続されることである。

【0037】

請求項19記載の発明の特徴は、請求項11記載の発明による文書情報処理装置であって、前記出力制約情報を設定する手段を有することである。

【0038】

請求項20記載の発明の特徴は、請求項11記載の発明による文書情報処理装置において、前記出力制約情報がネットワークを介して設定されることである。

【0039】

請求項21乃至30記載の発明は、複数の要素からなる文書情報の内容把握に必要な情報をできるだけ保存しつつ、全体の情報量を削減するための文書情報処理方法に係るもので、その特徴は以下の通りである。

【0040】

請求項21記載の発明の特徴は、文書情報の各要素の重要度を評価し、前記文書情報の各要素にその重要度に応じた操作を施すことにより、前記文書情報の情報量を削減することである。

【0041】

請求項22記載の発明の特徴は、請求項21記載の発明による文書情報処理方法において、前記文書情報中に記述された重要度定義情報に基づいて、前記文書情報の各要素の重要度を評価することである。

【0042】

請求項23記載の発明の特徴は、請求項21記載の発明による文書情報処理方法において、所定の重要度評価基準に基づいて、前記文書情報の各要素の重要度

を評価することである。

【0043】

請求項24記載の発明の特徴は、請求項21又は22記載の発明による文書情報処理方法において、特定の重要度レベルより低い重要度の要素を削除することである。

【0044】

請求項25記載の発明の特徴は、請求項24記載の発明による文書情報処理方法において、要素の属性によって前記特定の重要度レベルが異なることである。

【0045】

請求項26記載の発明の特徴は、請求項25記載の発明による文書情報処理方法において、テキスト属性の要素に関する前記特定の重要度レベルより、非テキスト属性の要素に関する前記特定の重要度レベルが高いことである。

【0046】

請求項27記載の発明の特徴は、請求項24記載の発明による文書情報処理方法において、テキスト属性の要素を削除対象から除外し、非テキスト属性の要素のみ削除対象とすることである。

【0047】

請求項28記載の発明の特徴は、請求項22又は23記載の発明による文書情報処理方法において、非テキスト属性の要素を、その重要度に応じた圧縮方法で圧縮することである。

【0048】

請求項29記載の発明の特徴は、請求項22又は23記載の発明による文書情報処理方法において、非テキスト属性の要素を、その重要度に応じた圧縮率で圧縮することである。

【0049】

請求項30記載の発明の特徴は、請求項22又は23記載の発明による文書情報処理方法において、第1の重要度レベルより重要度が低いテキスト属性の要素を削除し、第2の重要度レベルより重要度が低い非テキスト属性の要素を圧縮することである。

【0050】

請求項31乃至44記載の発明は、複数の要素からなる文書情報の内容把握に必要な情報をできるだけ保存しつつ、全体の情報量を削減するための文書情報処理装置に係るもので、その特徴は以下の通りである。

【0051】

請求項31記載の発明の特徴は、文書情報ソースより、複数の要素からなる文書情報を入力する文書情報入力手段と、前記文書情報入力手段による入力文書情報の各要素の重要度を評価する重要度評価手段と、前記入力文書情報の各要素に、前記重要度評価手段により評価された重要度に応じた操作を施すことにより前記文書情報の情報量を削減する文書情報処理手段と、前記文書情報処理手段により情報量を削減された文書情報を文書情報保存装置に出力する手段とを有することである。

【0052】

請求項32記載の発明の特徴は、請求項31記載の発明による文書情報処理装置において、前記重要度評価手段が、前記文書情報中に記述された重要度定義情報に基づいて前記文書情報の各要素の重要度を評価することである。

【0053】

請求項33記載の発明の特徴は、請求項31記載の発明による文書情報処理装置において、前記重要度評価手段が、所定の重要度評価基準に基づいて前記文書情報の各要素の重要度を評価することである。

【0054】

請求項34記載の発明の特徴は、請求項3.2又は33記載の発明による文書情報処理装置において、前記文書情報処理手段が、特定の重要度レベルより低い重要度の要素を削除することである。

【0055】

請求項35記載の発明の特徴は、請求項34記載の発明による文書情報処理装置において、要素の属性によって前記特定の重要度レベルが異なることである。

【0056】

請求項36記載の発明の特徴は、請求項35記載の発明による文書情報処理裝

置において、テキスト属性の要素に関する前記特定の重要度レベルより、非テキスト属性の要素に関する前記特定の重要度レベルが高いことである。

【0057】

請求項37記載の発明の特徴は、請求項34記載の発明による文書情報処理装置において、テキスト属性の要素を削除対象から除外し、非テキスト属性の要素のみ削除対象とすることである。

【0058】

請求項38記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置において、前記文書情報処理手段が、非テキスト属性の要素を、その重要度に応じた圧縮方法で圧縮することである。

【0059】

請求項39記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置において、前記文書情報処理手段が、非テキスト属性の要素を、その重要度に応じた圧縮率で圧縮することである。

【0060】

請求項40記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置において、前記文書情報処理手段が、第1の重要度レベルより重要度が低いテキスト属性の要素を削除し、第2の重要度レベルより重要度が低い非テキスト属性の要素を圧縮することである。

【0061】

請求項41記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置であって、前記文書情報ソースを有することである。

【0062】

請求項42記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置であって、前記文書情報ソースとネットワークを介して接続されることである。

【0063】

請求項43記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置であって、前記文書情報保存装置を有することである。

【0064】

請求項44記載の発明の特徴は、請求項32又は33記載の発明による文書情報処理装置であって、前記文書情報保存装置とネットワークを介して接続されることである。

【0065】

請求項45乃至48記載の発明はコンピュータが読み取り可能な記録媒体に係るもので、その特徴は以下の通りである。

【0066】

請求項45記載の発明の特徴は、請求項1乃至7のいずれか1項記載の発明による文書処理方法をコンピュータで実施するためのプログラムが記録されたことである。

【0067】

請求項46記載の発明の特徴は、請求項8乃至20のいずれか1項記載の発明による文書情報処理装置をコンピュータで実現するためのプログラムが記録されたことである。

【0068】

請求項47記載の発明の特徴は、請求項21乃至30のいずれか1項記載の発明による文書情報処理方法をコンピュータで実施するためのプログラムが記録されたことである。

【0069】

請求項48記載の発明の特徴は、請求項31乃至44のいずれか1項記載の発明による文書情報処理装置をコンピュータで実現するためのプログラムが記録されたことである。

【0070】

【発明の実施の形態】

以下、添付図面を参照し、本発明の実施の形態について説明する。

図1は本発明の一実施形態である文書情報処理装置の機能的ブロック構成の一例を説明するためのブロック図である。

【0071】

この文書情報処理装置100は、文書情報ソース111より、複数の要素からなる文書情報を入力するための文書情報入力部101を有する。ここでは、入力される文書情報は、後述するようなXMLで記述された構造化文書情報であるものとして説明する。文書情報ソース111は、通常は、文書情報が保存されている記憶装置であるが、XMLエディタのようなXML文書情報を作成するアプリケーション・ソフトも文書情報ソース111に含まれる。後述の文書情報保存装置115と文書情報ソース111とが共通の記憶装置である場合もあり得る。

### 【0072】

文書情報処理装置100はさらに、文書情報解析部102、文書情報処理部103、情報圧縮部104、及び作業用記憶部105を有する。文書情報解析部102は、XMLパーサと呼ばれる標準的なツールとして普及しているプログラムと同様な構文解析を入力文書情報に対して行うほか、入力文書情報の各要素の重要度の評価を行う手段である。文書情報処理部103は、入力文書情報の印刷、表示又は伝送のための文書（出力文書と呼ぶ）の生成や入力文書情報の情報量削減のための処理を行う手段である。情報圧縮部104は、入力文書情報中の静止画像情報、動画像情報又は音声情報の圧縮処理に利用されるものである。この情報圧縮部104は概念的には文書情報処理部103に含まれるものであるが、情報圧縮にハードウェアの各種符号化器が利用される可能性が高いいため、敢えて情報圧縮部104を文書情報処理部103から分離して示した。作業用記憶部105は、入力文書情報の記憶域や、文書情報解析部102、文書情報処理部103、情報圧縮部104などの処理に関連した情報などの一時記憶域を提供する手段である。

### 【0073】

文書情報処理装置100はさらに、文書情報処理部103により生成された出力文書（イメージデータ）を印刷装置112、表示装置113、ファクス装置114へそれぞれ出力するための出力部106、107、108、文書情報処理部103により情報量が削減された文書情報を文書情報保存装置115へ出力する出力部109、以上の各部（101～109）の動作の制御及び装置全体の動作の制御を司る制御部110を有する。ユーザ指示装置116は、ユーザが各種指

示や出力制約条件情報などを入力するための手段である。ユーザ指示装置116より入力された指示や情報の解釈や作業用記憶部105への保存は、制御部110によって行われる。

#### 【0074】

印刷装置112はプリンタなどの印刷専用装置に限定されるものではなく、イメージデータを入力として画像を形成するデジタル複写機なども包含するものである。表示装置113もディスプレイ装置のように画像表示専用の装置に限定されるものではなく、パソコンなどのイメージデータを入力として画像を表示できる機器なども包含するものである。また、ファクス装置114も、ファクスカード（ファクスモデム）などを備えるパソコン、その他の画像伝送装置を包含するものである。

#### 【0075】

この文書情報処理装置100は、全体を単一の装置として実現する形態、ネットワークで接続された複数の装置の集合体として実現する形態のいずれも採り得る。これについて図2を参照して説明する。

#### 【0076】

図2において、200はワークステーションなどの汎用のコンピュータであり、表示装置201、キーボードやマウスなどの入力装置202、ハードディスク装置203、印刷装置204などの周辺機器を備える。このコンピュータ200は、LAN（ローカルエリアネットワーク）210により印刷装置213、大容量記憶装置214、ファクス装置216、パソコンなどの他のコンピュータ215と接続され、さらにネットワーク間インターフェース手段211（ルータ、ゲートウェイなど）を介して他のネットワーク212（インターネットや他のLANなど）と接続される。

#### 【0077】

例えば、文書情報処理装置100を、コンピュータ200のハードウェア資源を利用してソフトウェアにより実現することができる。この場合、コンピュータ200の表示装置201を表示装置113として、印刷装置204を印刷装置112として、ハードディスク装置203又はコンピュータ200の内部メモリを

文書情報ソース111として、ハードディスク装置203を文書情報保存装置115として、入力装置202をユーザ指示装置116として、コンピュータ200の内蔵するファクスカード（不図示）をファクス装置114として、それぞれ利用することができる。このような集中形態の文書情報処理装置100又はその処理を実現するためのプログラムが記録された磁気ディスク、光ディスク、光磁気ディスク、半導体記憶素子などの記録媒体も本発明に包含される。

#### 【0078】

印刷装置112としてLAN210上の印刷装置213などを利用し、表示装置113及びユーザ指示装置116としてLAN210上のコンピュータ215やネットワーク212上のコンピュータ218の表示装置及び入力装置などを利用し、文書情報ソース111や文書情報保存装置115として、LAN210上的大容量記憶装置214やネットワーク212上のWWWサーバー217や他のコンピュータ218の記憶装置などを利用し、ファクス装置114としてLAN210上のファクス装置216などを利用することも可能である。このような分散形態の文書情報処理装置100又はその処理を、コンピュータ200上で実現するためのプログラムが記録された各種記録媒体も本発明に包含される。

#### 【0079】

このように、文書情報処理装置100は様々な実現形態をとり得るが、以下の説明においては専ら図1を参照する。

#### 【0080】

ここで、以下の説明に用いられる文書情報の例を図4及び図11に示す。いずれの例も、構造化文書記述言語であるXMLで記述された構造化文書情報である。図4の文書情報は、通常の処理をし適切なスタイルシートにより整形することにより、例えば図5に示すような2ページの文書として表示又は印刷される内容である。また、図11の文書情報は、通常の処理をし適切なスタイルシートにより整形することにより、例えば図12に示すような1ページの文書として表示又は印刷される内容である。

#### 【0081】

XMLで記述された文書情報は、XML宣言、文書型宣言、文書エンティティ

の3ブロックから構成される。図4の例では、1行目がXML宣言、3行目から17行目までが文書型宣言、19行目から44行目までが文書エンティティに相当する。図11の例では、1行目がXML宣言、3行目から19行目までが文書型宣言、21行目から39行目までが文書エンティティに相当する。なお、図4及び図11には行番号が記されているが、これは説明のために記したものであり、実際には必要ない。

#### 【0082】

XML宣言は、文書情報がXMLで記述された構造化文書情報であることを明示する。文書型宣言は、文書情報に含まれる要素や文書の論理構造といった文書型定義を記述する部分であり、要素の名前、要素間の親子関係、子要素の出現順序と出現回数などを規定する。文書エンティティでは、文書情報の内容を記述する。

#### 【0083】

文書情報を構成する要素には、タグと呼ばれる印付け（マークアップ）が行われる。山バーレン”<”、“>”で囲まれたdoc,title,mainなどの文字列がタグに相当する。このタグは、文書情報を構成する要素の意味を明確にする役割があり、文書情報の内容に応じて文書の作成者側で独自に定義する。

#### 【0084】

文書情報の論理構造は、最上位の要素が一つだけ存在し、その要素をルート（根）とよぶ。各要素は親となることができ、子の要素を含む構造になる。図4の例では、タグ「doc」がルートに相当し、その子として「title」、「main」、「misc」、「img」がある。このような階層的な親子関係を持つことから、構造化文書と呼ばれる。

#### 【0085】

文書エンティティは、文書情報の内容を記述しているが、ルートとなる要素の始まりを示すタグ（開始タグ）で始まり、終わりを示すタグ（終了タグ）で終わる。全ての要素は開始タグ、終了タグを持ち、各要素の内容は開始タグと終了タグの間に記述する。開始タグは<タグ名>、終了タグは</タグ名>と記述し、タグ名には文書型定義で定義した要素の名前を用いる。例えばタグ「doc」の開

始タグは「<doc>」、終了タグは「</doc>」である。

#### 【0086】

また、図4及び図11の例においては、文書情報を構成する各要素の重要度定義情報が文書情報中に記述されている。すなわち、文書型宣言ブロックにおいて、各要素の重要度レベル（level1～level3）が要素の属性として定義されている。図4の例では、例えば、タグ「title」の重要度レベルが

```
<!ATTLIST title viewpriority CDATA #FIXED "level1">
```

のように、viewpriority変数にlevel1を設定することによって重要度レベルが定義されている。level1, level2, level3はviewpriority変数の値であり、level1が最も高い重要度のレベルであり、level3が最も低い重要度のレベルである。同様の方法で、「sectiontitle」、「contents」、「misc」、「img」の各タグについても重要度レベルが定義されている。ただし、2レベル又は4レベル以上の重要度レベルを定義することも可能である。

#### 【0087】

ここでは、ATTLISTと呼ばれる属性情報として重要度レベルを定義しているが、ELEMENT、ENTITYなどにおいて重要度レベルを定義することも可能である。

#### 【0088】

このように文書情報中に各要素の重要度定義情報を記述する方法は、文書作成者側で各要素の重要度を柔軟かつ詳細に定義できる利点があり、また、文書処理装置100側では、文書情報中に記述された重要度定義情報から直接的に各要素の重要度を評価することができるという利点がある。しかし、文書情報中に各要素の重要度定義情報を必ずしも記述する必要はない。重要度定義情報が記述されていない場合には、後述するように、文書情報処理装置100において、文書情報の解析時に所定の重要度評価基準に従って各要素の重要度を評価する。

#### 【0089】

次に、文書処理管理システム100の全体的な動作について、図3に示すフローチャートに沿って説明する。

#### 【0090】

処理の対象となる文書情報や処理モードが、予めユーザによってユーザ指示装

置116を介し指示される。複数の文書情報ソース111が存在する場合には、その選択もユーザによって指定される。処理モードとしては、印刷装置112、表示装置113又はファクス装置114への出力文書を生成して出力するモードA1～A3と、文書情報をその情報量を削減して文書情報保存装置115へ出力するモードB1～B5を指定できる。処理モードA1～A3が指定された場合には、ユーザ指示装置116を介して出力先の指定も行われ、また、処理モードA2、A3の場合には出力制約条件情報（ページサイズとページ数）の設定も行われる。処理モードA1の場合には、出力制約条件情報は設定されず、デフォルトのページサイズが用いられるが、別のページサイズを指定することも可能である。処理モードB1～B5が指定される場合、複数の文書情報保存装置115が存在するならば、その選択指示も入力される。

#### 【0091】

ユーザが以上の指示操作の後に処理開始指示を入力すると、文書情報処理装置100においては、文書情報入力部101によって指定された文書情報を文書情報ソース111より入力する（ステップS1）。この入力文書情報は作業用記憶部105に格納される。

#### 【0092】

次に、文書情報解析部102において、まず入力文書情報のXML宣言ブロックを読み込み（ステップS2）、それがXML宣言であるならば（ステップS3、Yes）XML文書情報であると判断して次の処理ステップに進む。XML宣言でなければ（ステップS3、No）、対象外の文書と判断され、制御部110は処理を終了させる。

#### 【0093】

文書情報解析部102は、入力文書情報をXML文書情報である判断すると、入力文書情報の文書型宣言と文書エンティティを読み込み（ステップS4）、文書型定義すなわち文書型宣言で規定された規則を用いて、文書エンティティに記述されているタグと要素を分離し、構文解析木と呼ばれる木構造に変換する（ステップS5）。図4と図11の文書情報に関する木構造を図6と図13にそれぞれ模式的に示す。このような木構造は作業用記憶部105に格納される。このよ

うな構造解析の内容は、前述したように標準的なXMLパーサによる解析処理と基本的に同様である。

#### 【0094】

文書情報解析部102は、この構文解析の際に、各要素の重要度の評価も行い、その結果を構文解析木に記述する。例えば図6に見られるように、構文解析木の各節にはタグ、属性、要素の組が記述される。重要度に関する情報は対応する節の属性情報として管理される。例えば、「title」というタグの重要度を表すviewpriority属性の値はlevel1である。

#### 【0095】

図4と図11の例のように、文書情報中に各要素の重要度レベルが要素の属性情報として記述されているならば、その重要度レベルをそのまま要素の重要度として評価する。しかし、入力文書情報中に要素の重要度定義情報が記述されていない場合には、所定の重要度評価基準に従って各要素の重要度を評価する。この重要度評価基準の例を挙げれば、テキスト属性の要素の重要度を非テキスト属性（静止画像、動画像、音声など）の要素の重要度より高く評価し、テキスト属性の要素の中では、文字列長（文字数）が短いほど重要度を高く評価する。文書情報の内容把握のためには、一般に、テキスト属性の要素が重要であり、その中でもタイトルなどの文字列長の短い要素が重要度が高いからである。ただし、これに限定されるものではない。非テキスト属性の要素の中で、静止画像の要素の重要度を動画像や音声などの要素より高く評価するなどの基準を定めることも可能である。また、いくつかの重要度評価基準を予め定めておき、ユーザに重要度評価基準を選択させることも可能である。

#### 【0096】

構文解析及び重要度評価を終了すると、制御部110はユーザにより指定された処理モードに応じて処理の流れを切り替える（ステップS6）。

#### 【0097】

処理モードA1、A2又はA3が指定された場合には、文書情報処理部103で出力文書（イメージ）を生成する処理を実行させ（ステップS103）、生成された出力文書をユーザが指定した出力先へ出力部106、107又は108を

介して出力させ（ステップS8）、処理を終了する。なお、出力先としてファクス装置113が指定された場合には、送信先のファクス番号もユーザにより予め指定される。

#### 【0098】

処理モードB1～B5のいずれかが指定された場合には、制御部110は、文書情報処理部103で入力文書情報の情報量を削減する処理を実行させ（ステップS9）、情報量削減後の文書情報（構造化文書情報）をユーザに指定された文書情報保存装置115へ出力部109を介し出力させ（ステップS10）、処理を終了する。情報量削減処理で画像情報などの情報圧縮が必要となる場合には、情報圧縮部104が利用される。

#### 【0099】

以下、各処理モードにおける文書情報処理部103の処理（ステップS7又はS9）の内容について説明する。

#### 【0100】

##### 《処理モードA1の説明》

この処理モードにおける処理内容を図7に示すフローチャートに沿って説明する。ただし、図4の文書情報が入力されたものとする。

#### 【0101】

図6の構文解析木全体をルートのタグ「doc」より左から右へ深さ優先で辿り、未選択の要素の中で重要度の最も高い要素を選択する（ステップS20）。最初に選択される要素は、最高の重要度レベルlevel1である「タイトル」の要素であり、この要素の内容すなわち文字列「タイトル」を予め作業用記憶部105に用意された出力文書（最初は空白状態）の1ページ目の先頭より配置する（ステップS22）。構文解析木の「タイトル」要素には選択済みのフラグ情報を付記する。

#### 【0102】

再び構文解析木を同様に辿るが、重要度レベルlevel1の未選択の要素はないため、その次に高い重要度レベルlevel2の「1章のタイトル」の要素を選択し、その内容である文字列「1章のタイトル」を出力文書に続けて配置する。この

要素に選択済みのタグ情報を付記する。

【0103】

次に重要度レベルlevel2の「2章のタイトル」の要素を選択し、その内容である文字列「2章のタイトル」を出力文書に続けて配置する。

【0104】

このようにして重要度の高い要素から順に選択し、その内容を出力文書上に配置する。最後に「imagefile.gif」の要素を選択し、その内容である画像情報を出力文書に配置すると、未選択の要素が見つからなくなるため（ステップS21, Yes）、処理は終了する。

【0105】

なお、作成日付の要素と作成者名の要素の重要度レベルは、タグ「misc」の属性値として記述された重要度レベルlevel2が用いられる。

【0106】

ここでは、要素を1つずつ選択しながら、選択した要素を出力文書に配置したが、要素選択処理と出力文書への要素配置処理とを分離し、全ての要素の選択を行った後に、要素を選択された順に出力文書に配置するようにしてもよい。この場合、出力文書への要素配置処理において要素の選択順を認識できるようにするため、要素選択処理において、例えば、構文解析木上で選択した要素にその選択順位を表すフラグ情報を付記する。このような構成の方法及び装置も本発明に包含される。

【0107】

このようにして重要度の高い要素から順に配置された出力文書が生成される。この出力文書の全ページを印刷装置112へ出力し印刷させた場合、印刷用紙の節約を期待することはできないが、1ページ目には例えば図8に示すように文書情報の重要度の高い要素が含まれるため、1ページ目に目を通すだけで入力文書情報の内容を確認することができる。1ページ目だけでは十分な内容把握ができない場合には2ページ目以降にも目を通すことになるが、2ページ目以降でも重要な要素から順に配置されているため、内容確認を効率的に行うことができる。また、印刷ページの指定を行って例えば1ページ目だけを印刷させることにより

、印刷用紙を節約することもできる。

【0108】

出力文書をFAX装置114に出力してファクス送信させる場合も、同様の効果を期待できる。また、出力文書を表示装置113へ出力して表示させた場合にも、重要度の高い要素から優先して表示されるため、最終ページまでスクロールすることなく、文書情報の内容を迅速に確認することができる。

【0109】

《処理モードA2の説明》

この処理モードにおける処理内容を図9に示すフローチャートに沿って説明する。ここでも、図4の文書情報が入力されたものとする。

【0110】

まず、処理モードA1の場合と同様に、図6の構文解析木全体をルートのタグ「doc」より左から右へ深さ優先で辿り、未選択の要素の中で重要度の最も高い要素を選択する（ステップS30）。最初に選択される要素は、最高の重要度レベルlevel1である「タイトル」の要素であり、この要素に選択済みであることと選択順位が1番であることを示すフラグ情報を付記する。

【0111】

1つの要素を選択すると、その要素を出力文書に配置した場合に占有するスペースを計算し（ステップS32）、その占有スペースを、これまでに選択された要素の占有スペースの積算値に加算する（ステップS33）。得られた積算スペースと、出力制約条件情報により指定された出力文書のページサイズ及びページ数で決まるスペース上限との比較判定を行う（ステップS34）。

【0112】

積算スペースがスペース上限以下ならば（ステップS34, No）、再び構文解析木全体を辿って、未選択要素の中で重要度が最も高い要素を選択し（ステップS30）、その要素の占有スペースを計算し（ステップS32）、その占有スペースをこれまでの積算スペースに加算し（ステップS33）、得られた積算スペースとスペース上限との比較判定を行う（ステップS34）。

【0113】

選択した要素の占有スペースの積算スペースがスペース上限を越えると（ステップS34, Yes）、最後に選択した要素のフラグ情報を未選択に変更することにより、その要素を選択された要素群より削除する（ステップS35）。つまり、最後に選択された要素は出力文書から削除される結果となる。次に、構文解析木を参照して、選択済み要素をその選択順の早いものから順に選び、その内容を、作業用記憶部105に用意した出力文書に1ページ目の先頭から配置することにより、重要度の高い順に選択された要素のみを、その順に配置した出力文書を生成する（ステップS36）。積算スペースがスペース上限を越える以前に全ての要素が選択された場合には、直ちにステップS36に進むため、全ての要素が選択された順に出力文書上に配置されることになる。

#### 【0114】

なお、ステップS32で、選択した要素を出力文書に実際に配置し、ステップS35で、出力文書に最後に配置した要素を出力文書より削除するようにしてもよい。当然、ステップS36は不要になる。このような構成の方法及び装置も本発明に包含される。

#### 【0115】

出力制約条件情報で例えば出力文書のページサイズがA4サイズ、ページ数が1ページに指定された場合、例えば図8に示すような内容の1ページの出力文書が生成される。この出力文書を印刷装置112に出力し印刷させるならば、文書情報の内容の効率的把握が可能になるとともに、文書情報をそのまま印刷する場合よりも印刷用紙を節約できる。

#### 【0116】

出力文書をファクス装置114に出力しファクス送信させた場合も、受信側で文書情報の内容を効率的に把握できるとともに記録紙を節約でき、さらにファクス送信時間も短縮できる。出力文書を表示装置113に出力し表示させた場合にも、画面のスクロール操作を殆ど行うことなく、文書情報の内容を素早く確認することができる。出力文書をネットワークを介して印刷装置112、表示装置113又はファクス装置114へ出力する場合には、その伝送時間も短縮できる。

#### 【0117】

### 《処理モードA3の説明》

この処理モードにおける処理内容を図10に示すフローチャートに沿って説明する。ここでも、図4の文書情報が入力されたものとする。

#### 【0118】

処理モードA2の場合と同様な要素選択（ステップS40）、占有スペースの計算（ステップS42）、占有スペースの積算（ステップS43）、積算スペースとスペース上限との比較判定（ステップS44）を繰り返す。

#### 【0119】

積算スペースがスペース上限を越えると（ステップS44, Yes）、選択した要素のサイズを縮小する処理を行う（ステップS45）。このサイズ縮小処理は、選択された全ての要素を対象として行うことも、一部の要素、例えば最後に選択された1要素又は数要素を対象として行うことも可能である。処理負担を軽くするために後者が一般に有利である。なお、サイズ縮小処理の対象範囲をユーザが選択できるようにしてもよい。テキスト要素のサイズ縮小は、フォントサイズを例えば1ポイント下げるることによって行う。画像の要素のサイズ縮小は、画像サイズを例えば1割だけ小さくするように解像度を調整することによって行う。

#### 【0120】

次に、このようなサイズ縮小処理後の積算スペースを再計算し（ステップS46）、算出した積算スペースとスペース上限との比較判定を行う（ステップS47）。積算スペースがスペース上限を越える場合には、サイズをさらに縮小し（ステップS45）、積算スペースを再計算し（ステップS46）、積算スペースとスペース上限との比較判定を行う（ステップS47）。

#### 【0121】

このようにして積算スペースがスペース上限以下まで減少すると、選択された要素を選択順に出力文書に配置する（ステップS48）。この際、サイズ縮小処理の対象となった要素は、サイズ縮小後のフォントサイズ又は画像サイズで出力文書に配置される。したがって、本処理モードA3によれば、処理モードA2に比べ同じページサイズ、同じページ数の出力文書に1つだけ要素を余分に配置で

きる。

#### 【0122】

積算スペースがスペース上限を越える前に未選択の要素が見つからなくなったら場合には（ステップS41, Yes）、ステップS45～ステップS47はスキップされ、全ての要素が重要度の高いものから順に、元のサイズのまま出力文書に配置される。

#### 【0123】

なお、例えばステップS42で、選択した要素を出力文書に実際に配置し、例えばステップS46で、サイズ縮小後の要素を出力文書に再配置するようにしてもよい。当然、ステップS48は不要になる。このような構成の方法及び装置も本発明に包含される。

#### 【0124】

本処理モードA3においても、処理モードA2の場合と同様に、出力文書を印刷装置112やファクス装置113に出力するならば、文書情報の内容の効率的な確認と印刷用紙又は記録紙の節約が可能となり、また出力時間やファクス送信時間を短縮できる。出力文書を表示装置113に出力し表示させれば、文書情報の内容を素速く確認できるようになる。

#### 【0125】

##### 《処理モードB1の説明》

この処理モードにおける処理内容を図14に示すフローチャートに沿って説明する。ただし、図11の文書情報が入力されたものとする。

#### 【0126】

図13の構文解析木をルートのタグ「doc」より左から右へ深さ優先で1回だけ通り、文書情報の要素を1つずつ順に選択する（ステップS50）。

#### 【0127】

選択した要素の重要度が最高の重要度レベルlevel1であるか調べ（ステップS51）、level1ならば、その要素を、それがテキスト属性の要素であるか非テキスト属性であるかを問わず、そのまま残す。

#### 【0128】

ステップS50で選択した要素の重要度がlevel1でなければ(level2又はlevel3のいずれかである)、その重要度が1つ低い重要度レベルlevel2であるか調べる(ステップS52)。重要度がlevel2でなれば、つまり最低の重要度レベルlevel3であるということがあるので、その要素を削除する(ステップS53)。

#### 【0129】

重要度がlevel2ならば、その要素がテキスト属性であるか判定する(ステップS54)。テキスト属性の要素ならば、そのまま残すが、画像などの非テキスト属性の要素ならば削除する(ステップS55)。

#### 【0130】

以上の処理を文書情報の各要素について繰り返し、最後の要素まで処理すると(ステップS56, Yes)、処理を終了する。

#### 【0131】

以上の説明から明らかなように、本処理モードB1においては、文書情報の内容把握のために一般的に重要度が高いテキスト属性の要素に関しては、重要度がlevel2までの要素を残し、それより重要度が低い要素だけを削除し、テキスト属性の要素に比べ一般的に重要度が低い画像や音声などの非テキスト属性の要素に関しては、重要度がlevel1の要素だけを残し、それより重要度が低い要素を削除する。このようにして、文書情報の内容把握に必要な情報をできるだけ保存しつつ、全体の情報量を削減することができる。

#### 【0132】

図11の文書情報を本処理モードB1で処理して情報量を削減したものを表示又は印刷すると、図15に示すような内容になる。タグ「contents」とタグ「img1」には重要度level3が設定されているので、その要素が文書情報から削除されたことを確認できる。

#### 【0133】

このようにして情報量が削減された文書情報(構造化文書情報)を文書情報保存装置115に出力し保存するならば、元の文書情報をそのまま保存する場合に比べ、保存のために必要な記憶容量を節減できる。また、文書情報保存装置11

5がネットワークを介して接続されている場合には、文書情報の伝送時間を短縮できる。また、文書情報保存装置115に保存された文書情報は、元の文書情報の内容確認に必要な情報を保存しているため、それを印刷又は表示することによって、元の文書情報の内容を支障なく確認できるばかりでなく、元の文書情報の重要度の低い要素が削除されているため、より効率的な内容確認が可能になる。

## 【0134】

## 《処理モードB2の説明》

この処理モードにおける処理フローを図16に示す。図16のフローチャートと、図14のフローチャートとを対照すれば容易に理解できるように、本処理モードB2と処理モードB1との相違点は、処理モードB1の要素削除ステップS55が圧縮処理ステップS65に置き換えられていることであり、これ以外のステップS60～S64、S66は処理モードB1の対応ステップS50～S54、S56と同じ内容の処理ステップである。

## 【0135】

すなわち、重要度がlevel2の非テキスト属性の要素は、処理モードB1では削除されるが、本処理モードB2ではステップS65で情報圧縮処理が施される。図11の文書情報の場合、タグ「img2」の要素「imagefile2.gif」が情報圧縮処理の対象となる。

## 【0136】

ステップS65の情報圧縮処理は情報圧縮部104を利用して実行される。静止画像の情報圧縮処理には、離散コサイン変換と呼ばれる圧縮方法や、ハフマン符号化と呼ばれる圧縮方法など、公知の圧縮方法を利用できる。このような圧縮方法に関する公知文献としては、「データ圧縮ハンドブック」(1994年3月31日発行、トッパン、M. ネルソン原著)などがある。また、動画像の情報圧縮にはMPEG(Moving Picture Experts Group)、音声の情報圧縮にはMP3(MPEG1 Audio Layer 3)などの公知の圧縮方法を利用できる。なお、画像については、上に述べたような圧縮方法に変えて、サイズを縮小したり、解像度を下げたり、量子化間隔を大きくするなどの方法で情報量を圧縮することも可能である。

## 【0137】

以上の説明から明らかなように、本処理モードB2は、画像や音声などの非テキスト属性の要素に関しては、重要度がlevel1の要素はそのまま残すが、重要度がlevel2の要素は情報圧縮を施してから保存することが特徴である。文書情報の内容把握のために一般的に重要なテキスト属性の要素に関しては、処理モードB1と同様に、重要度がlevel2までの要素を残し、それより重要度が低い要素だけを削除する。

## 【0138】

## 《処理モードB3の説明》

この処理モードにおける処理内容を図17に示すフローチャートに沿って説明する。ただし、図11の文書情報が入力されたものとする。

## 【0139】

図13の構文解析木をルートのタグ「doc」より左から右へ深さ優先で1回だけ通り、文書情報の要素を1つずつ順に選択する（ステップS70）。

## 【0140】

選択した要素がテキスト属性の要素であるか調べ（ステップS71）、テキスト属性ならばそのまま残す。選択した要素が非テキスト属性の要素であるならば、その重要度がlevel1であるか調べ（ステップS72）、level1ならばそのまま残すが、level1以外ならば削除する（ステップS73）。

## 【0141】

以上の処理を文書情報の各要素について繰り返し、最後の要素まで処理すると（ステップS73、Yes）、処理を終了する。

## 【0142】

このように、本処理モードB3は、文書情報の内容を把握するために一般的に重要なテキスト属性の要素は、その重要度に関わらず全て残すことが特徴である。テキスト属性の要素に比べ一般的に重要度が低い画像や音声などの非テキスト属性の要素に関しては、処理モードB1と同様に重要度がlevel1の要素だけを残し、それより重要度が低い要素は削除する。このようにして、文書情報の内容把握に必要な情報をできるだけ保存しつつ、全体の情報量を削減することができ

る。

【0143】

《処理モードB4の説明》

この処理モードにおける処理内容を図18に示すフローチャートに沿って説明する。ただし、図11の文書情報が入力されたものとする。

【0144】

図13の構文解析木をルートのタグ「doc」より左から右へ深さ優先で1回だけ通り、文書情報の要素を1つずつ順に選択する（ステップS80）。

【0145】

選択した要素がテキスト属性の要素であるか調べ（ステップS81）、テキスト属性ならばそのまま残す。選択した要素が非テキスト属性の要素であるならば、その重要度がlevel1であるか調べ（ステップS82）、level1ならば情報圧縮処理モードとして低圧縮モードを、それ以外の重要度レベル（level2又はlevel3）ならば高圧縮モードを選択し（ステップS83、S84）、選択したモードで情報圧縮処理を施す（ステップS85）。以上の処理を文書情報の各要素について繰り返し、最後の要素まで処理すると（ステップS86、Yes）、処理を終了する。

【0146】

図11の文書情報では、タグ「img2」の要素「imagefile2.gif」は重要度がlevel1であるため低圧縮モードで情報圧縮されるが、タグ「img1」の要素「imagefile1.gif」は重要度がlevel3であるため高圧縮モードで情報圧縮される。

【0147】

ここで、低圧縮モードの圧縮処理とは、情報量の削減効果は限られるが画像などの再生品質の劣化が少ない低圧縮率の圧縮処理である。高圧縮モードの圧縮処理とは、情報量を大幅に削減できるが画像などの再生品質の劣化が著しい高圧縮率の圧縮処理である。

【0148】

それぞれのモードの圧縮処理に同じ圧縮方法を用いてもよいが、モードによって異なる圧縮方法を切り替えて使用してもよい。例えば、静止画像の要素に関し

て、低圧縮モードの圧縮処理に前述のハフマン符号化を利用し、高压縮モードの圧縮処理に前述の離散コサイン変換を利用することができる。このような圧縮方法によらず、サイズ、解像度、量子化間隔などを重要度に応じて調整することにより、同様の情報量削減を行ってもよい。

#### 【0149】

以上に述べたように、本処理モードB4は、文書情報の内容を把握するために一般的に重要なテキスト属性の要素については、処理モードB3と同様に重要度に関わらず全て残す。テキスト属性の要素に比べ一般的に重要度が低い非テキスト属性の要素に関しては、重要度がlevel1の要素を、その再生品質（画質もしくは音質）の悪化の少ない低圧縮モードで情報圧縮し、それより重要度が低い要素を再生品質の劣化が大きいが情報量削減効果の大きい高压縮モードで情報圧縮することが特徴である。このようにして、文書情報の内容把握に必要な情報をできるだけ保存しつつ、全体の情報量を削減することができる。

#### 【0150】

##### 《処理モードB5の説明》

この処理モードにおける処理フローを図19に示す。図19のフローチャートと、図18のフローチャートとを対照すれば容易に理解できるように、本処理モードB5と処理モードB4との相違点は、ステップS92で重要度がlevel1でないと判定した場合にはステップS94でlevel2であるか判定し、level1とlevel2の要素に対しては低圧縮モードと高压縮モードをそれぞれ選択して圧縮処理を施すが（ステップS93、S95、S97）、level3の要素は削除する（ステップS96）ことである。ステップS90～S93、S95、S97、S98は、処理モードB4の対応ステップS80～S86と同じ内容の処理ステップである。

#### 【0151】

なお、低圧縮モードと高压縮モードの中間的な圧縮率のモードを用意し、ステップS94でlevel2と判定された要素に対して、その中間的な圧縮率のモードを選択して圧縮処理を施し、ステップS94でlevel2でない（level3である）と判定された要素に対して高压縮モードを選択して圧縮処理を施すようにしても

よい。このような構成の方法及び装置も本発明に包含される。

【0152】

以上の説明においては、重要度は3レベルに評価されたが、4レベル以上の重要度評価を行い、重要度レベルによる要素の操作もしくは処理の内容を、より詳細に制御してもよいことは明らかである。

【0153】

【発明の効果】

請求項1乃至20記載の発明によれば、XMLで記述された構造化文書情報のような、複数の要素からなる文書情報の内容を効率的に把握可能な文書を、画像出力装置により印刷又は表示させ、あるいは画像伝送装置により送信し受信側で記録させることができる。画像出力装置や画像伝送装置で消費される印刷用紙又は記録紙を節約することができる。画像伝送装置の文書送信時間を短縮できる。画像出力装置や画像伝送装置が文書情報処理装置とネットワークを経由して接続される場合には、画像出力装置や画像伝送装置への文書伝送に要する時間を短縮できる、等々の効果を得られる。

【0154】

請求項21乃至44記載の発明によれば、XMLで記述された構造化文書情報のような、複数の要素からなる文書情報の内容把握に必要な情報を保存しつつ、その全体の情報量を削減することができる。したがって、文書情報の保存に必要な記憶容量を節減できる。また、文書情報の保存のための装置が文書情報処理装置とネットワークを介して接続されている場合には、文書情報の伝送時間を短縮できる。また、文書情報保存装置に保存された文書情報を印刷又は表示することにより、その内容を支障なく把握することができるとともに、元の文書情報の重要度の低い要素が削除されているため、より効率的に内容を確認できる、等々の効果を得られる。

【0155】

請求項45乃至48記載の発明によれば、一般的なコンピュータを利用し請求項1乃至44記載の発明を容易に実施できるという効果を得られる。

【図面の簡単な説明】

【図1】

本発明による文書情報処理装置の機能的ブロック構成の一例を示すブロック図である。

【図2】

文書情報処理装置の実現形態を説明するための図である。

【図3】

文書情報処理装置の全体的な動作を説明するためのフローチャートである。

【図4】

XMLで記述された構造化文書情報の一例を示す図である。

【図5】

図4の文書情報を普通に処理し適切なスタイルシートで整形した出力文書の一例を示す図である。

【図6】

図4の文書情報の解析により得られる構文解析木を示す図である。

【図7】

処理モードA1の処理内容を説明するためのフローチャートである。

【図8】

処理モードA1により図4の文書情報から生成される出力文書の1ページ目の内容を例示する図である。

【図9】

処理モードA2の処理内容を説明するためのフローチャートである。

【図10】

処理モードA3の処理内容を説明するためのフローチャートである。

【図11】

XMLで記述された構造化文書情報の一例を示す図である。

【図12】

図11の文書情報を普通に処理し適切なスタイルシートで整形した出力文書の一例を示す図である。

【図13】

図11の文書情報の解析により得られる構文解析木を示す図である。

【図14】

処理モードB1の処理内容を説明するためのフローチャートである。

【図15】

図11の文書情報を処理モードB1により処理した結果を出力して得られる文書の内容を示す図である。

【図16】

処理モードB2の処理内容を説明するためのフローチャートである。

【図17】

処理モードB3の処理内容を説明するためのフローチャートである。

【図18】

処理モードB4の処理内容を説明するためのフローチャートである。

【図19】

処理モードB5の処理内容を説明するためのフローチャートである。

【符号の説明】

100 文書情報処理装置

101 文書情報入力部

102 文書情報解析部

103 文書情報処理部

104 情報圧縮部

105 作業用記憶部

106~109 出力部

110 制御部

111 文書情報ソース

112 印刷装置

113 表示装置

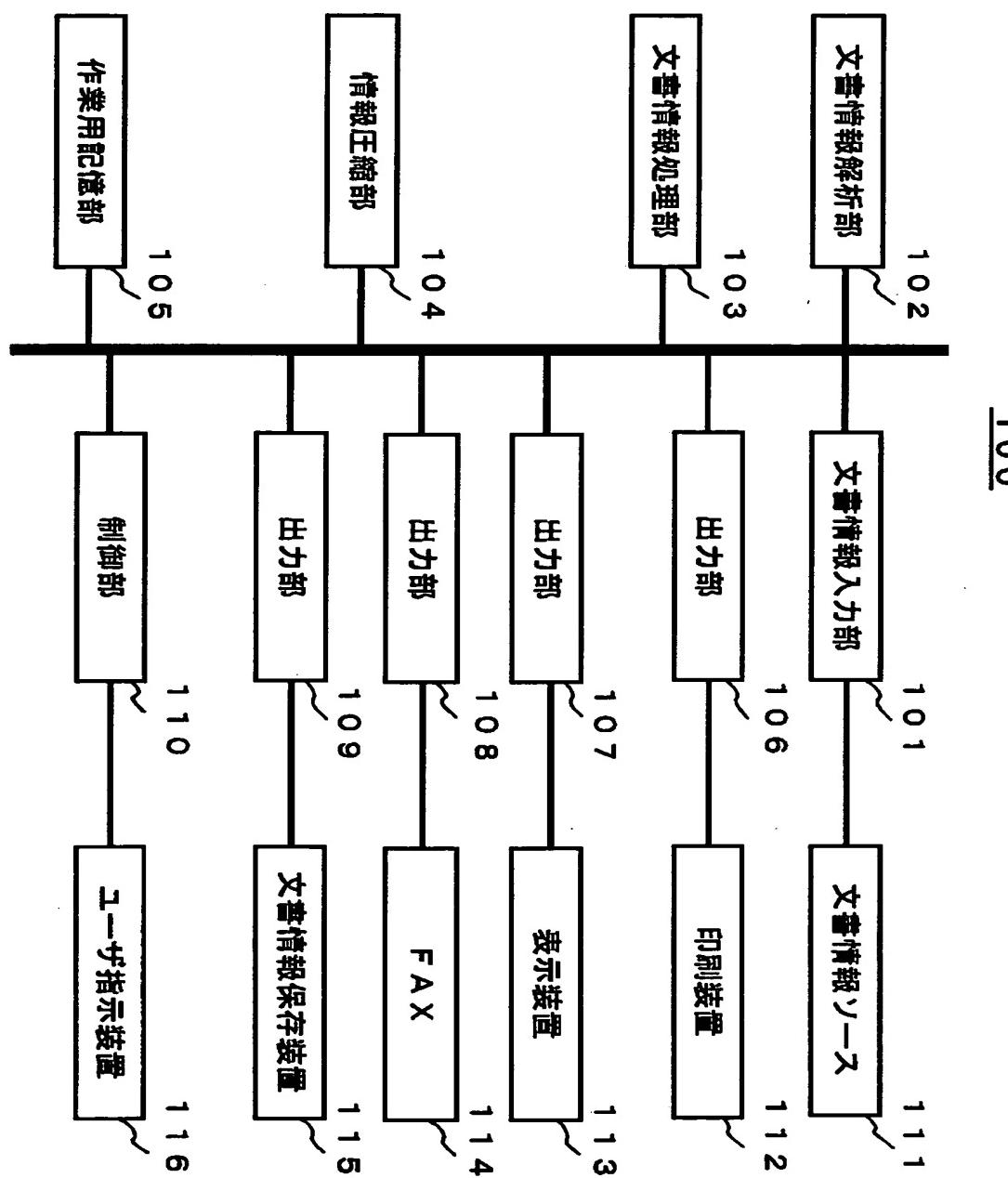
114 ファクス装置

115 文書情報保存部

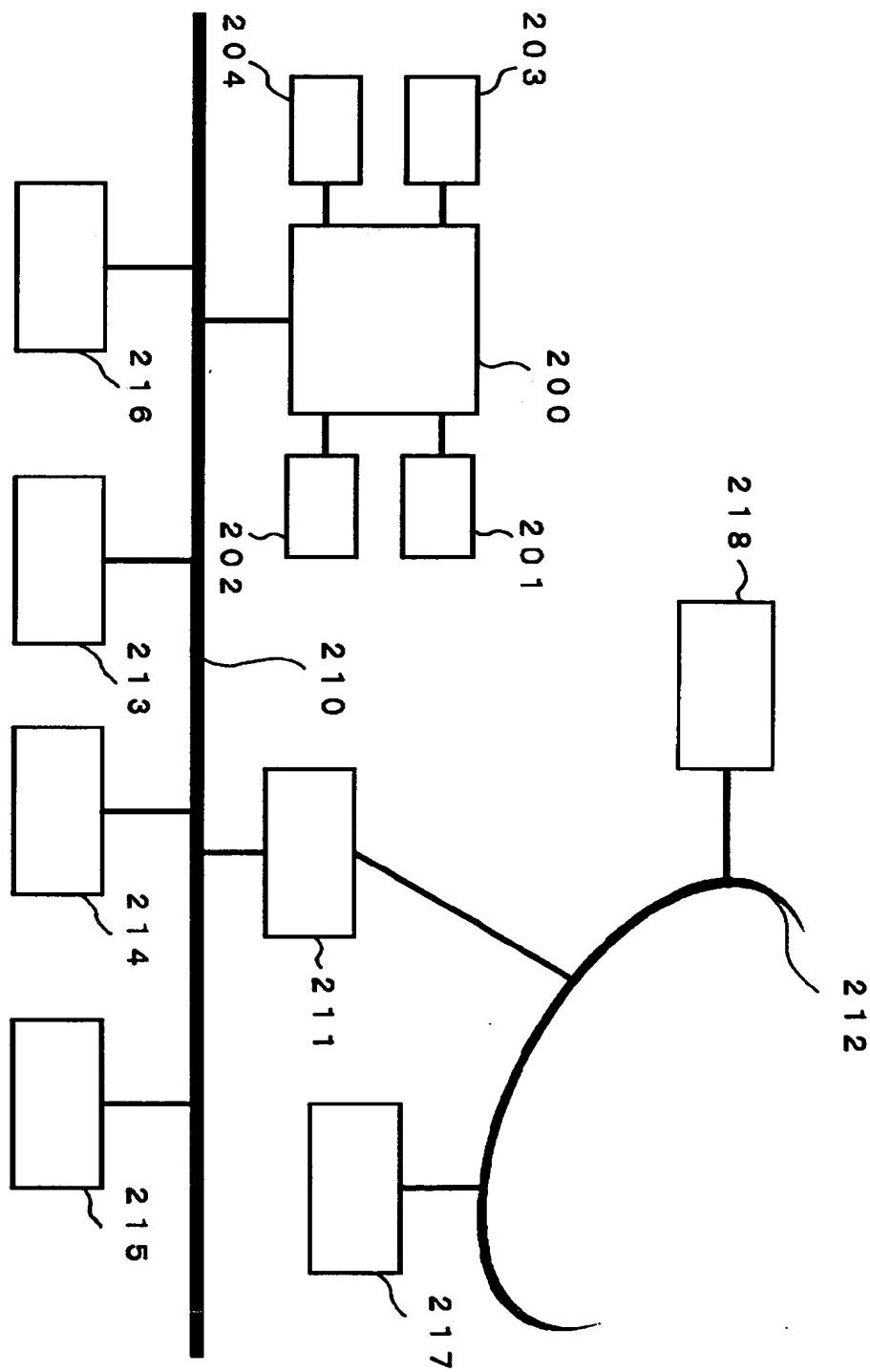
116 ユーザ指示装置

【書類名】図面

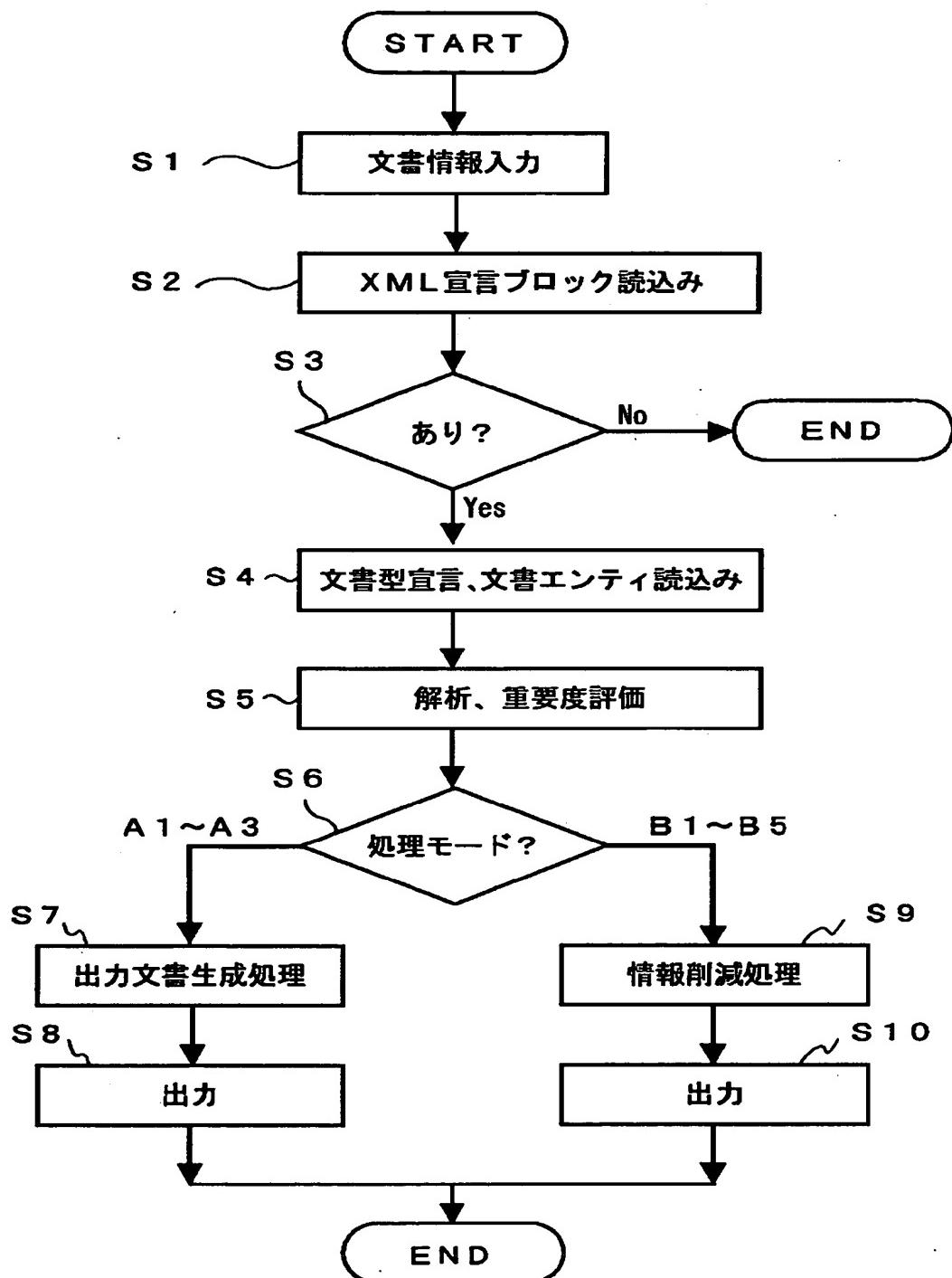
【図1】



【図2】



【図3】



【図4】

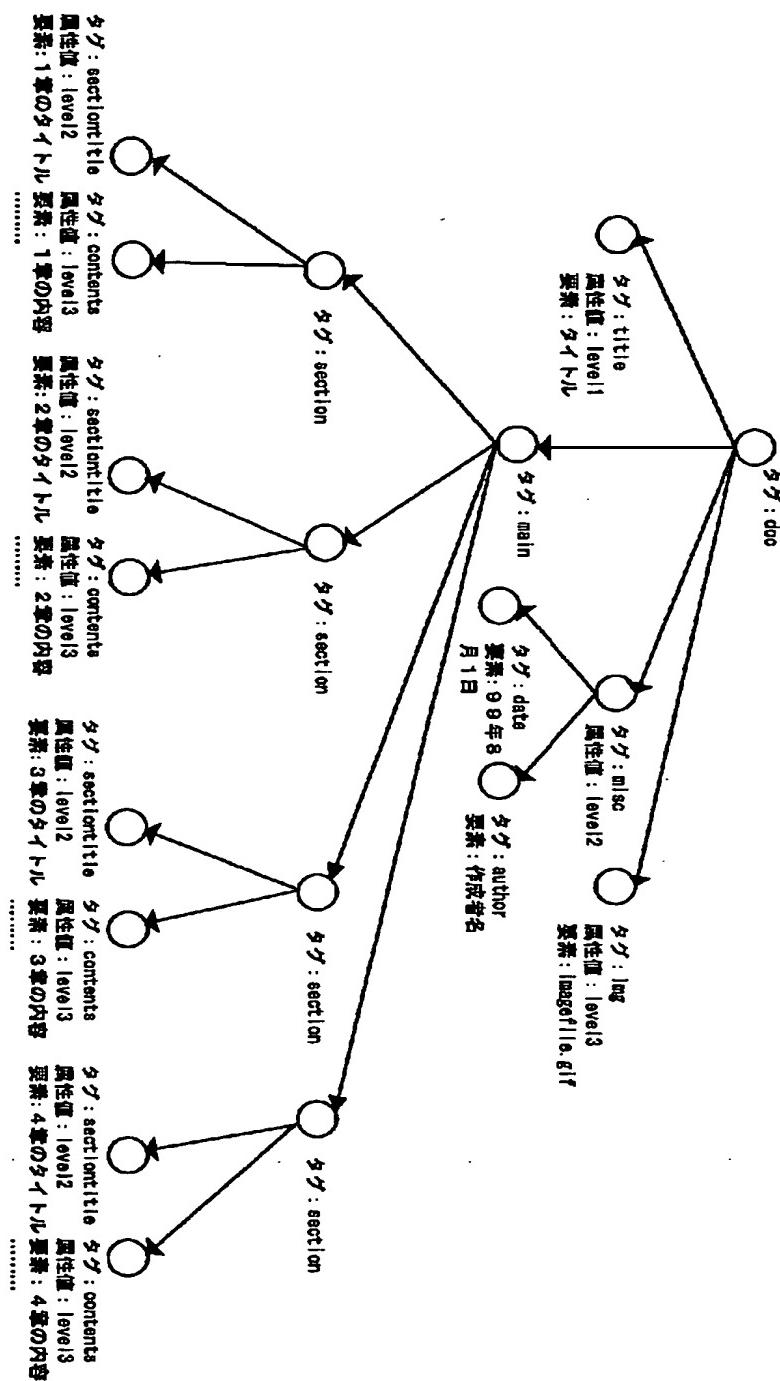
```
1 : <?xml version="1.0" encoding="Shift_JIS"?>
2 :
3 : <!ELEMENT doc (title,main,misc,img)>
4 : <!ELEMENT title (#PCDATA)>
5 : <!ATTLIST title viewpriority CDATA #FIXED "level1">
6 : <!ELEMENT main (section*)>
7 : <!ELEMENT section (sectiontitle,contents)>
8 : <!ELEMENT sectiontitle (#PCDATA)>
9 : <!ATTLIST sectiontitle viewpriority CDATA #FIXED "level2">
10 : <!ELEMENT contents (#PCDATA)>
11 : <!ATTLIST contents viewpriority CDATA #FIXED "level3">
12 : <!ELEMENT misc (date,author)>
13 : <!ATTLIST misc viewpriority CDATA #FIXED "level2">
14 : <!ELEMENT date (#PCDATA)>
15 : <!ELEMENT author (#PCDATA)>
16 : <!ELEMENT img EMPTY>
17 : <!ATTLIST img viewpriority CDATA #FIXED "level3">
18 :
19 : <doc>
20 : <title>タイトル</title>
21 : <main>
22 : <section>
23 : <sectiontitle>1章のタイトル</sectiontitle>
24 : <contents>1章の内容。.....</contents>
25 : </section>
26 : <section>
27 : <sectiontitle>2章のタイトル</sectiontitle>
28 : <contents>2章の内容。.....</contents>
29 : </section>
30 : <section>
31 : <sectiontitle>3章のタイトル</sectiontitle>
32 : <contents>3章の内容。.....</contents>
33 : </section>
34 : <section>
35 : <sectiontitle>4章のタイトル</sectiontitle>
36 : <contents>4章の内容。.....</contents>
37 : </section>
38 : </main>
39 : <misc>
40 : <date>99年8月1日</date>
41 : <author>作成者名</author>
42 : </misc>
43 : </img>
44 : </doc>
```

【図5】

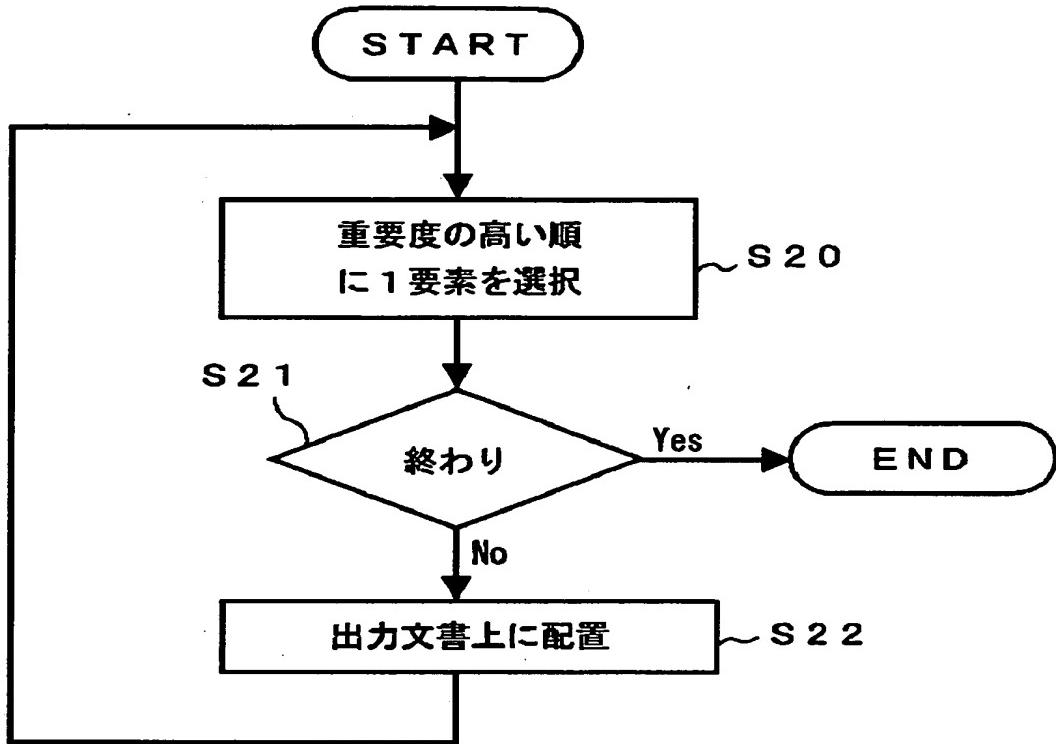
タイトル  
1章のタイトル  
1章の内容。  
.....  
  
2章のタイトル  
2章の内容。  
.....  
  
3章のタイトル  
3章の内容  
.....

4章のタイトル  
4章の内容。  
.....  
  
99年8月1日  
作成者名  
  

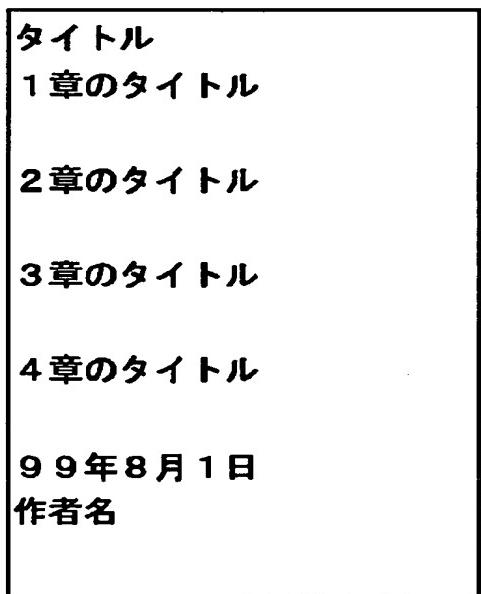

【図6】



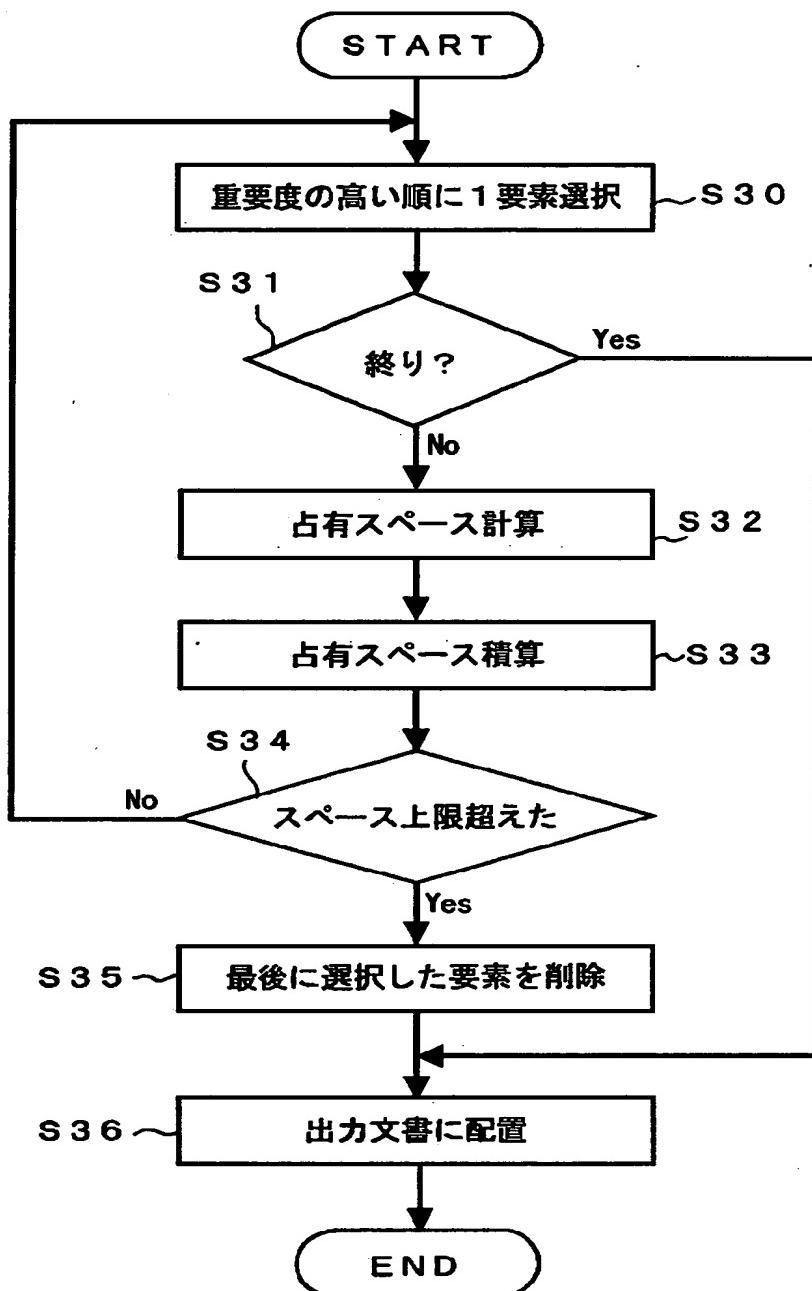
【図7】



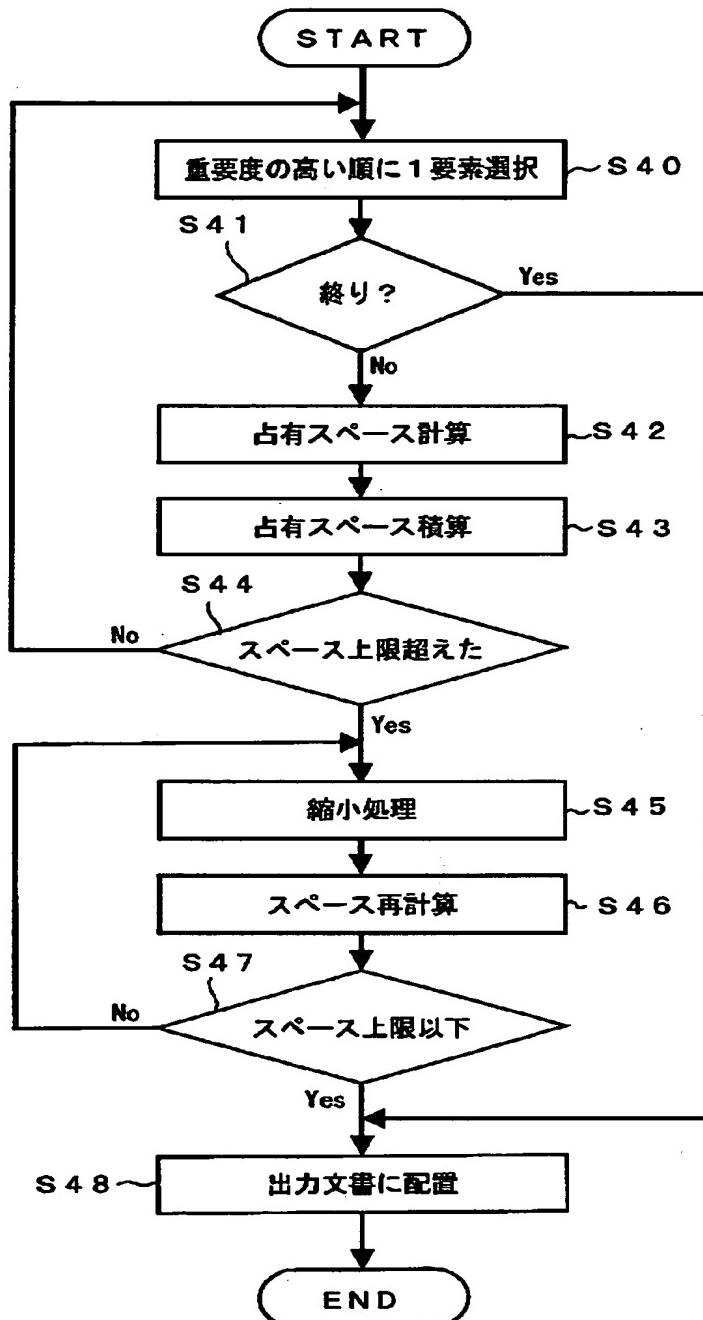
【図8】



【図9】



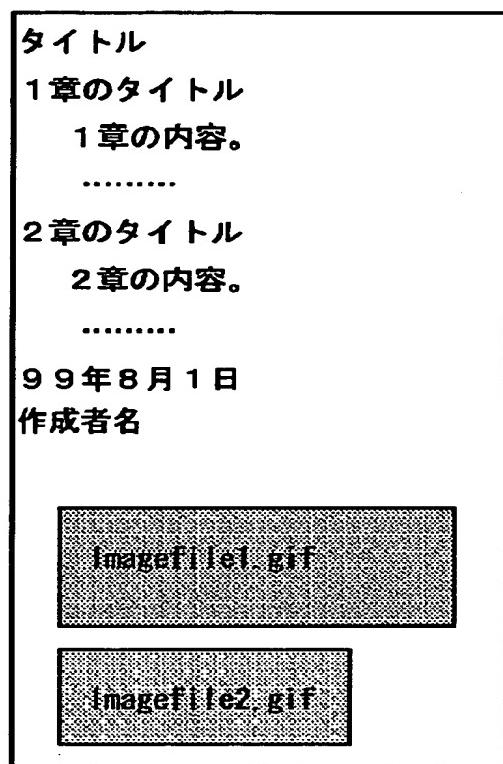
【図10】



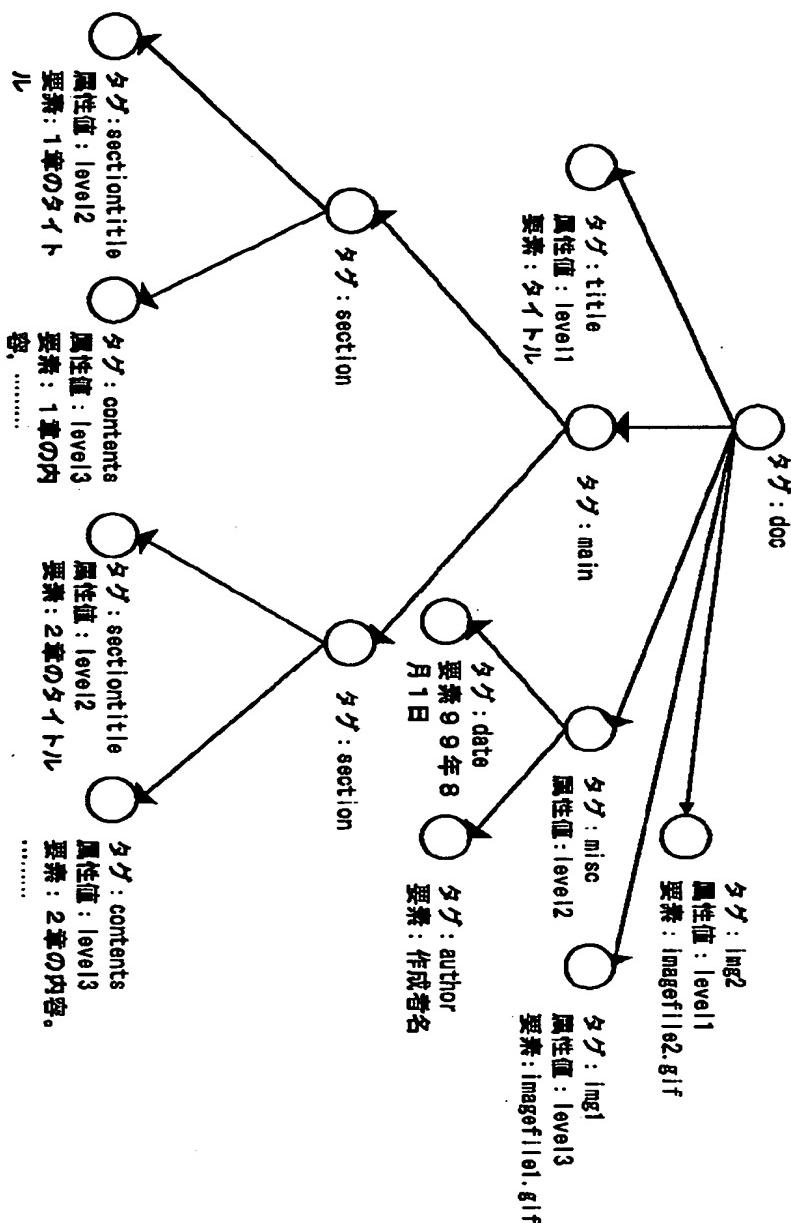
## 【図11】

```
1 : <?xml version="1.0" encoding="Shift_JIS"?>
2 :
3 : <!ELEMENT doc (title,main,misc,img1,img2)>
4 : <!ELEMENT title (#PCDATA)>
5 : <!ATTLIST title viewpriority CDATA #FIXED "level1">
6 : <!ELEMENT main (section*)>
7 : <!ELEMENT section (sectiontitle,contents)>
8 : <!ELEMENT sectiontitle (#PCDATA)>
9 : <!ATTLIST sectiontitle viewpriority CDATA #FIXED "level2">
10 : <!ELEMENT contents (#PCDATA)>
11 : <!ATTLIST contents viewpriority CDATA #FIXED "level3">
12 : <!ELEMENT misc (date,author)>
13 : <!ATTLIST misc viewpriority CDATA #FIXED "level2">
14 : <!ELEMENT date (#PCDATA)>
15 : <!ELEMENT author (#PCDATA)>
16 : <!ELEMENT img1 EMPTY>
17 : <!ATTLIST img1 viewpriority CDATA #FIXED "level3">
18 : <!ELEMENT img2 EMPTY>
19 : <!ATTLIST img2 viewpriority CDATA #FIXED "level1">
20 :
21 : <doc>
22 : <title>タイトル</title>
23 : <main>
24 : <section>
25 : <sectiontitle>1章のタイトル</sectiontitle>
26 : <contents>1章の内容。………</contents>
27 : </section>
28 : <section>
29 : <sectiontitle>2章のタイトル</sectiontitle>
30 : <contents>2章の内容。………</contents>
31 : </section>
32 : </main>
33 : <misc>
34 : <date>99年8月1日</date>
35 : <author>作成者名</author>
36 : </misc>
37 : <img1 src="imagefile1.gif"></img1>
38 : <img2 src="imagefile2.gif"></img2>
39 : </doc>
```

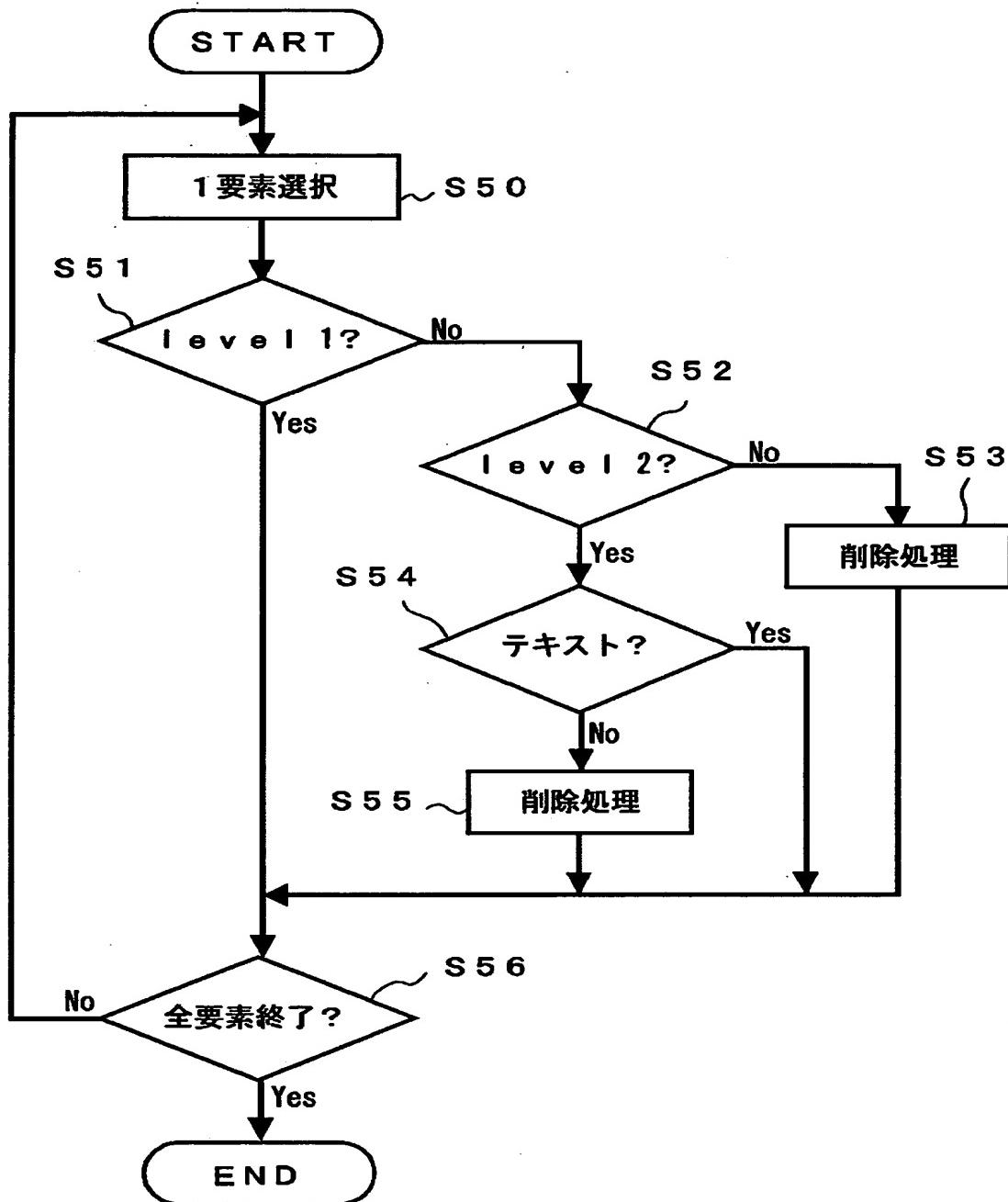
【図12】



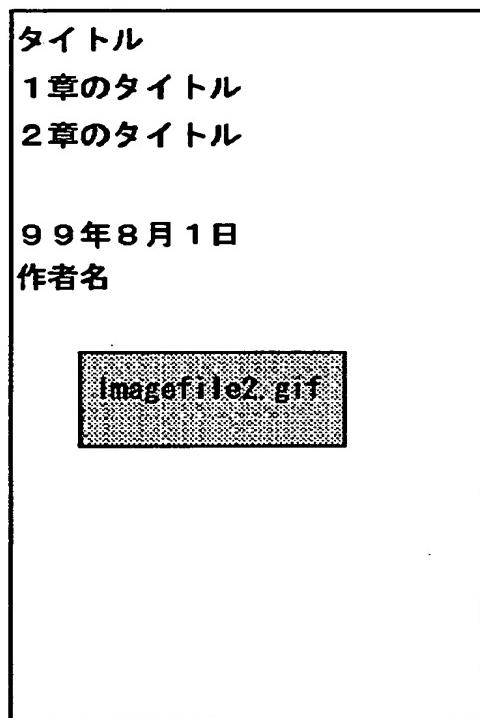
【図13】



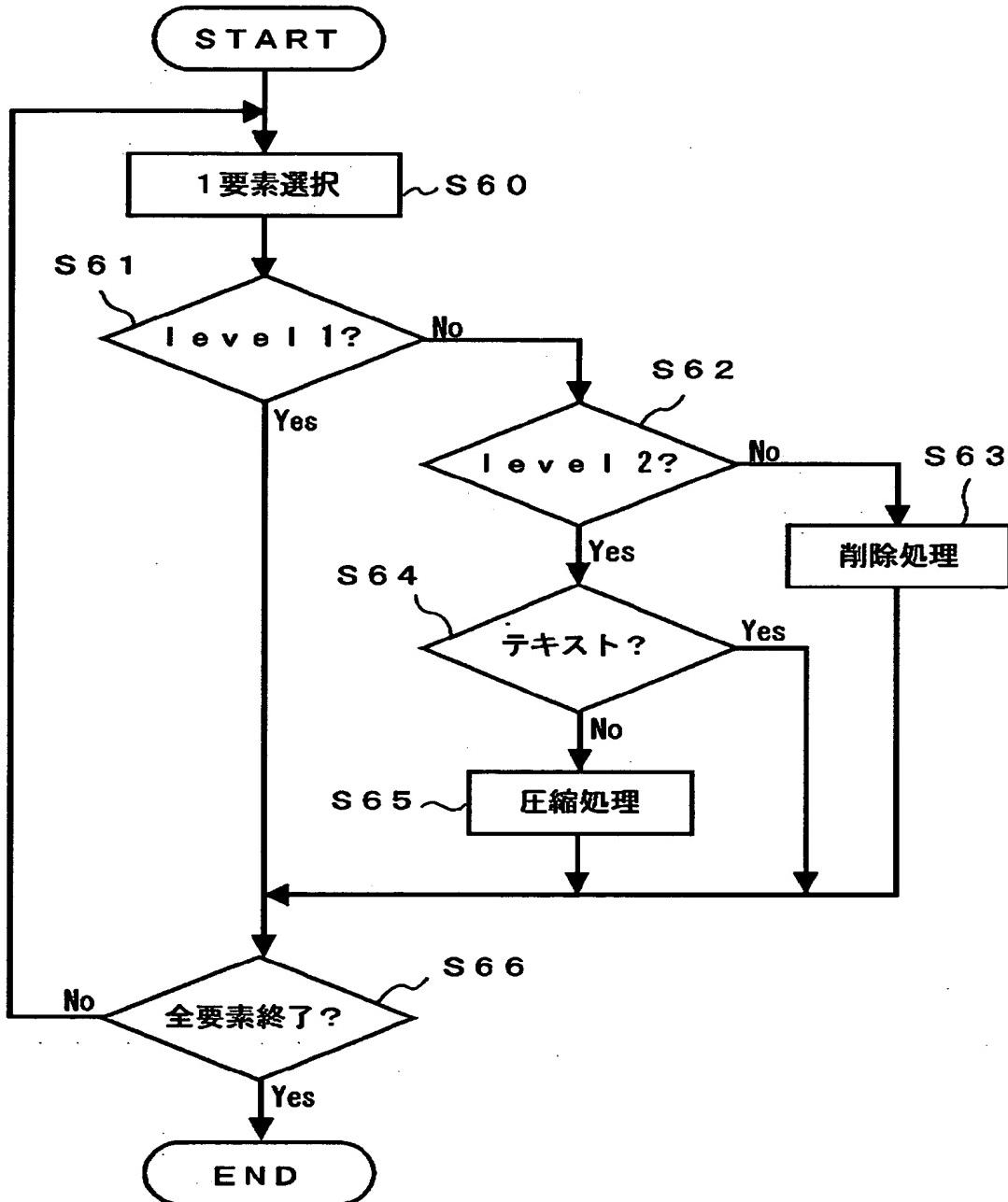
【図14】



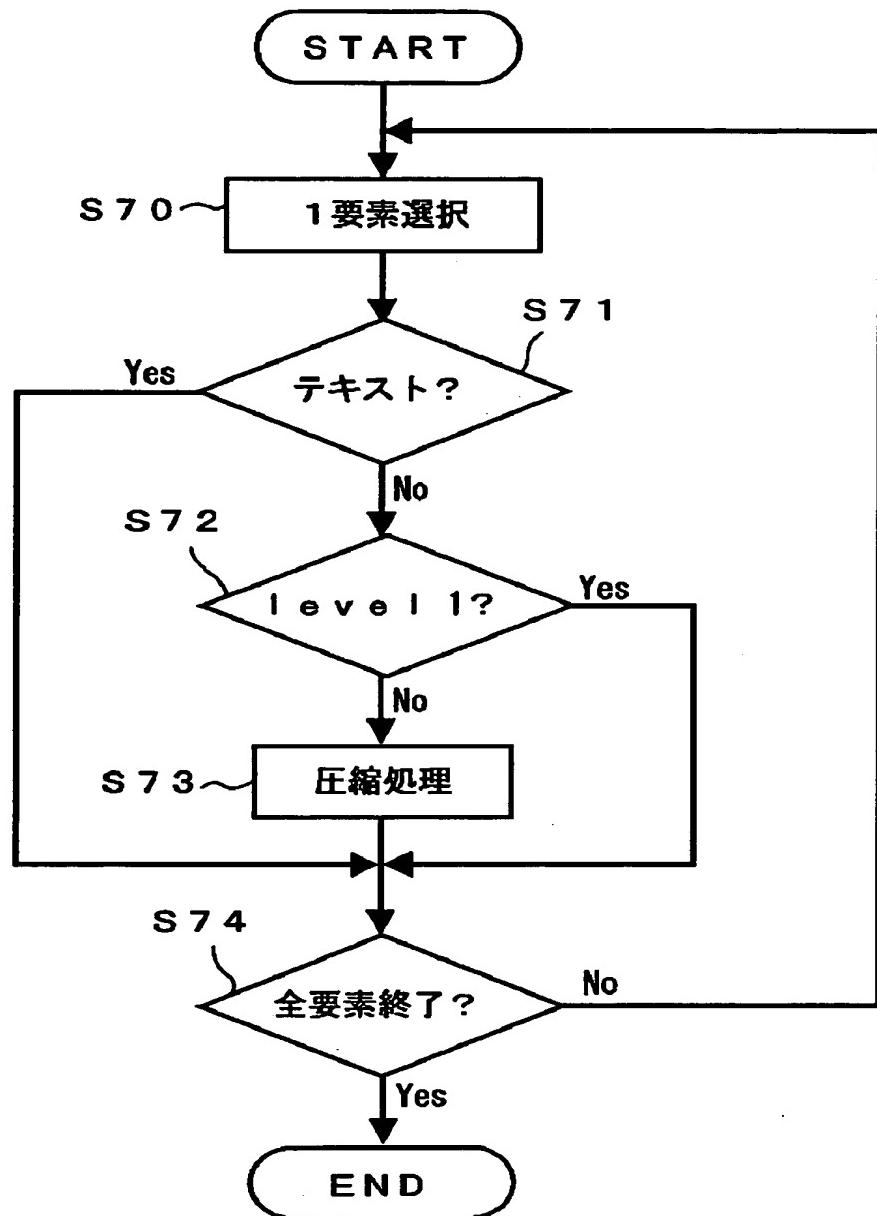
【図15】



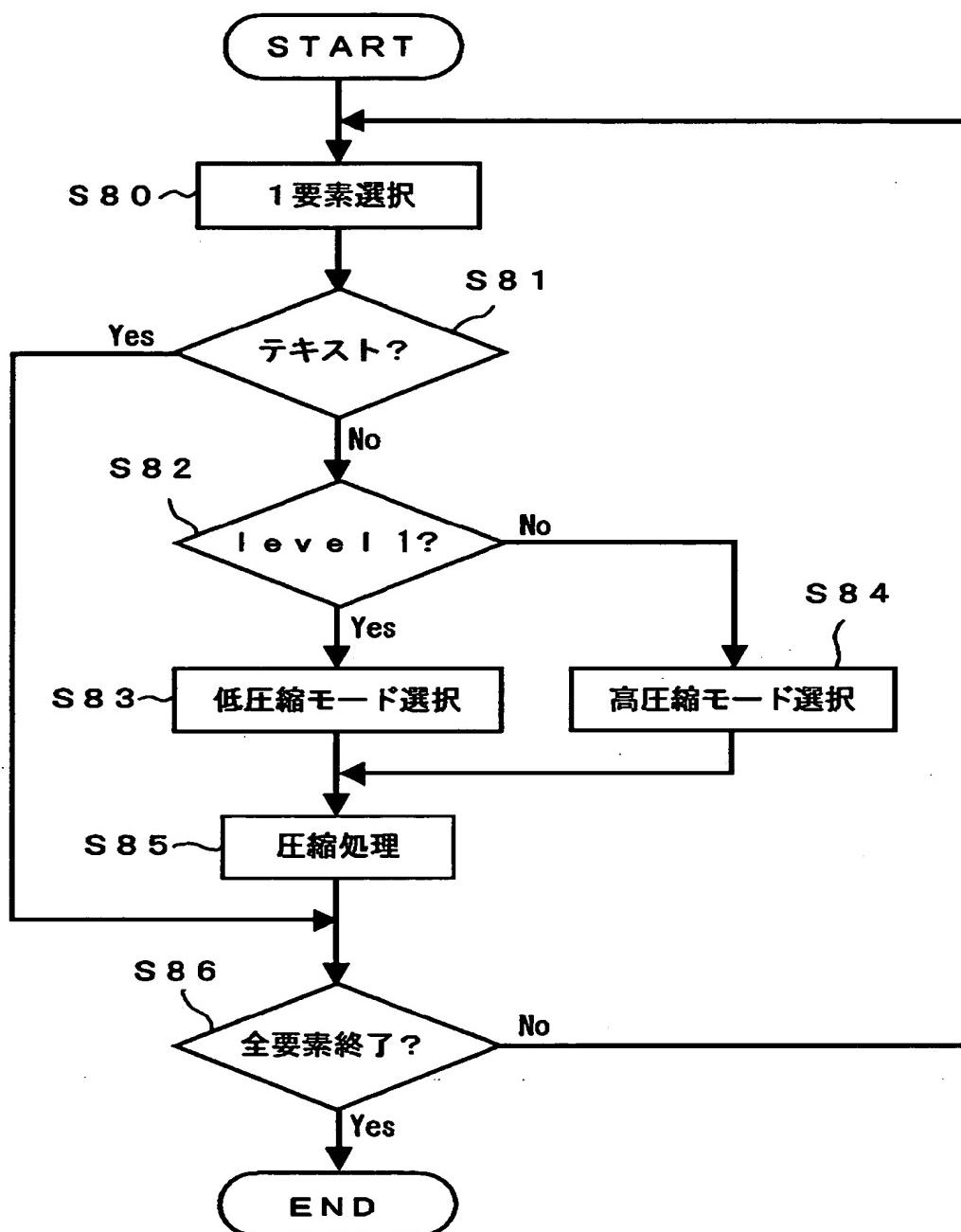
【図16】



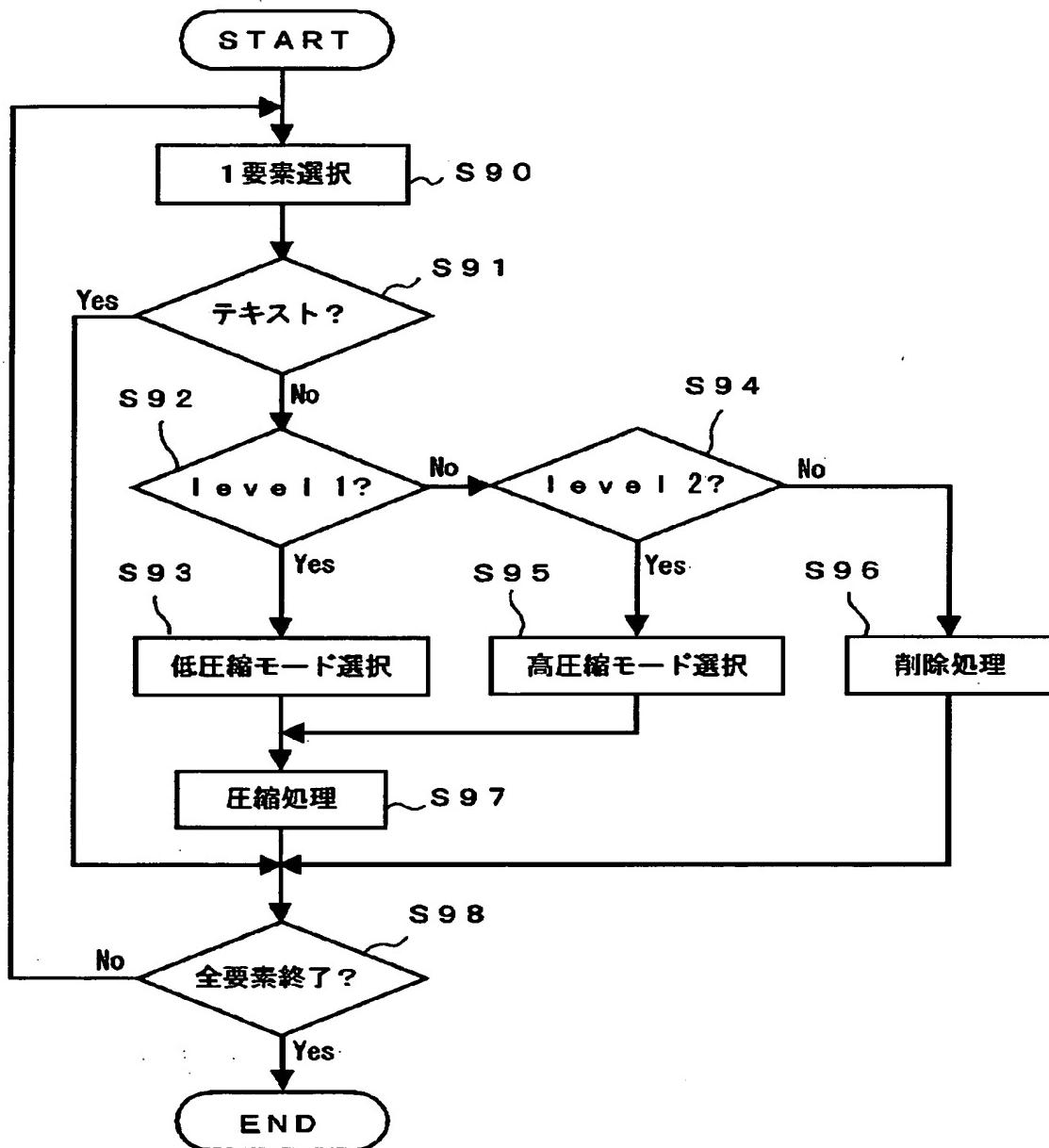
【図17】



【図18】



【図19】



【書類名】 要約書

【要約】

【課題】 文書情報を印刷又は表示する場合に、文書情報の内容の効率的な把握と印刷用紙の節約を可能にする。文書情報の保存のための記憶容量を減らす。

【解決手段】 文書情報入力部101により入力された文書情報の構文解析と、文書情報の各要素の重要度評価が文書情報解析部102で行われる。処理モードA1～A3が指定された場合、文書情報処理部103で、文書情報の要素を重要度の高いものから順に選択して出力文書に配置する。この出力文書は印刷装置112などに出力される。処理モードB1～B5が指定された場合、文書情報処理部103で文書情報の情報量削減処理が行われる。処理後の文書情報は文書情報保存部115に出力される。

【選択図】 図1

特2000-214394

出願人履歴情報

識別番号 [000006747]

1. 変更年月日 1990年 8月24日

[変更理由] 新規登録

住 所 東京都大田区中馬込1丁目3番6号

氏 名 株式会社リコー